Resilient Infrastructure and Building Security

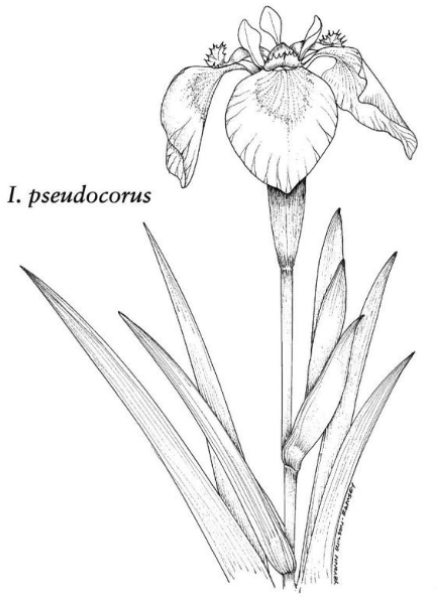
Ingwar, Mads Ingerslew

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

RESILIENT INFRASTRUCTURE AND BUILDING SECURITY

MADS INGERSLEW INGWAR



PERSISTENT AUTHENTICATION FOR LOCATION-BASED SERVICES

KONGENS LYNGBY – PHD-2014-322

Acknowledgments

I would like to express my gratitude to the many people whose presence and assistance have been invaluable in the last few years. A special thanks to my supervisor Christian Damsgaard Jensen, whose encouragement, expertise, and understanding were tremendous and without whom I'd be none the wiser, not least in the area of French pop culture. This thesis have been a truly multi-disciplinary effort, contingent on the expertise provided by all the parters in the RIBS project, to whom credit is due for their help and engagement. I would also like to thank my friends and colleagues in the department. Thanks to Naveed for his invaluable help and statistical advice at times of need. Thanks to Karin for all the small things and for making the bureaucracy bearable. Thanks to Michael for taking me out for that beer and sticking around afterwards. A special thanks to my family for the support and understanding through this entire process called my life. Finally, my very special thanks to Sidsel, without whose love, encouragement and tireless editing assistance, I would not have finished this thesis.

TECHNICAL UNIVERSITY OF DENMARK
DEPARTMENT OF APPLIED MATHEMATICS AND COMPUTER SCIENCE
BUILDING 303B, DK-2800 KGS. LYNGBY, DENMARK
PHONE +45 4525 3031, COMPUTE@COMPUTE.DTU.DK
WWW.COMPUTE.DTU.DK
PHD-2014-322, ISSN 0909-3192

Find yourself a girl and settle down
Live a simple life in a quiet town
Steady as she goes

— The Raconteurs

SUMMARY

Traditional authentication systems are considered persistent as they rarely limit the time the authentication is valid. Conversely, sensor-based authentication systems are considered transient as they allow continuous authentication of the users.

In this thesis we present *Persistent Authentication for Location-based Services*, as a new approach to authentication that combines traditional access control systems with the sensing technologies and tracking capabilities offered by smart environments. Persistent authentication enables the secure provision of location-based services through non-intrusive authentication of mobile users in a smart environment. The objective is to shift the current authentication paradigm from a single discrete event to a continuous session. This is accomplished by utilising the contextual awareness provided by the smart environments to track principals from the point of initial authentication to the point where authorisation is requested by the location-based services.

Facial recognition and appearance analysis are integrated in the persistent authentication system as remote biometric experts that operate at a distance and require no interaction from the users. The experts perform continuous authentication by processing samples of the biometric modalities as they become available.

Combining scores from multiple biometric experts is known as sensor fusion. A common challenge in this field is that the results from evaluating different biometric characteristic are usually incompatible, as they have different score ranges as well as different probability distributions. *Error-rate-based fusion* is presented as a novel fusion technique that transforms individual scores, from different biometric systems, into objective evidences and combine them using Bayesian inference.

Persistent authentication offers an effective integrated protection measure that is distributed directly in the facility and is non-intrusive to the public and affordable to the facility owners. Persistent authentication is suitable for security sensitive applications and can help protect the facility against insiders, intruders and hostile reconnaissance.

RESUMÉ

Traditionelle autentificeringssystemer anses som vedvarende da de sjældent begrænser den tid autentificeringen er gyldig. Omvendt, betragtes sensorbaserede autentificeringssystemer som transiente, da de tillader kontinuerlig autentificering af brugerne.

I denne afhandling præsenterer vi *Persistent Authentication for Location-based Services*, som ny tilgang til autentificering der kombinere traditionelle adgangskontrolsystemer med sensor teknologi og sporing i smarte miljøer. Persistent Authentication muliggøre sikre lokationsbaserede tjenester gennem ikke-invasiv autentificering af mobile brugere i smarte miljøer. Formålet er at ændre det nuværende autentificeringsparadigme fra en diskret begivenhed til en kontinuerlig session. Dette opnås ved at udnytte de kontekstuelle faktorer fra det smarte miljø til at spore brugerene fra deres indledende autentificering til det punkt hvor autentificeringen skal bruges af de lokationsbaserede tjenester.

Biometriske eksperter der udfører ansigtsgenkendelse og analyse af udseende er integreret i systemet. Disse eksperter opererer på afstand og kræver ingen interaktion fra brugerne. Eksperterne foretager løbende autentificering ved at processere prøver af de biometriske modaliteter som de er tilgængelig.

Kombinationen af scorer fra flere biometriske eksperter er kendt som sensor fusion. En udfordring i sensor fusion er at resultaterne fra evalueringen af de forskellige biometriske modaliteter kan være uforenelige, da de har forskellige score intervaller og sandsynlighedsfordelinger. *Error-based-fusion* præsenteres som en ny fusions metode der omdanner de individuelle score fra forskellige biometriske systemer til objektive beviser der kombineres ved hjælp af Bayesiansk interferens.

Persistent Authentication er en effektiv integreret beskyttelsesforanstaltning der distribueres direkte i bygningerne. Persistent Authentication er ikke-invasiv for brugerne og økonomisk overkommelig for bygningsejerne. Persistent Authentication er tiltænkt sikkerhedsfølsomme anvendelser og kan hjælpe med at beskytte bygninger mod eksterne og interne trusler.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PUBLICATIONS

PEER-REVIEWED PUBLICATIONS

M. Ingwar, C. Jensen: "Persistent Authentication for Location-based Services". *Re-submitted with minor revisions, Computers and Security*, 2014.

M. Ingwar, C. Jensen: "Remote Biometrics for Robust Persistent Authentication". *In Proceedings of the 6th SETOP International Workshop on Autonomous and Spontaneous Security, in conjunction with the 18th annual European research event in Computer Security (ESORICS)*, 2013.

M. Ingwar, N. Ahmed, C. Jensen: "Error-Rate-based Fusion of Biometric Experts". *In Proceedings of the 11th Annual Conference on Privacy, Security and Trust (PST)*, 2013.

M. Ingwar, C. Jensen: "Towards Secure Intelligent Buildings". *In Proceedings of the 5th Nordic Workshop on Dependability and Security (NODES)*, 2011.

OTHER PUBLICATIONS

M. Ingwar, I. Petropouleas: "White Paper on Protection Measurement Requirements", *Resilient Infrastructure and Building Security Deliverable 7.1*, 2013

M. Ingwar, T. Nissen: "Context Analysis and Ethical Framework Report", *Resilient Infrastructure and Building Security Deliverable 2.1*, 2011

Part I

SECURE BUILDINGS

1

INTRODUCTION

As the presence of technology progressively increases to pervade our urban environment it is envisaged that the advancements and protection measures employed to secure buildings, too, will become truly pervasive. In the editor's foreword to Homeland Security [1], Reiter et al. state that for secure buildings, the primary advancements considered are pervasive technologies that involve the use of surveillance, access control, and data analysis to detect patterns of activities that presage attacks. These technologies help prevent unauthorised entry and access and play a major role in preventing, detecting, and providing early warning of attacks.

This thesis focuses on two of the main research topics in pervasive computing: analysing sensor-based person tracking technologies and identifying anomalous behaviour using a given set of sensors; both person tracking and anomaly detection are large and active research areas that will be investigated in the context of detecting offenders and disrupting their attack plans. The contributions in this thesis are to implement and integrate the appropriate technologies from these areas to form an effective situational protection measure, which will be evaluated with respect to the given context.

State of the art protection measures that are developed and implemented in the absence of contextual information are of limited effectiveness. In security analysis, humans are identified as both the most important and the weakest point in the implementation of security. Thus, there is a need for a new approach that integrates human and organisational factors at the design stage, and merges security requirements with the complex constraints of inhabitable environments.

With regard to building security, human threats can be divided into three groups:

- *Illegitimate access*. Attacks conducted by individuals who gain unauthorised entry into restricted areas, either through infiltration, tailgating or forced entry.

- *Insiders*. Attacks conducted by individuals who belong to the target organisation or have other affiliations that grant them authorisation to access restricted areas.
- *Hostile reconnaissance*. Scouting of the organisation prior to an attack, which is an integral part of the operational activity for offenders.

Buildings can be analysed as designed artefacts that gives rise to specific patterns of human interaction and behaviour, such as varying movement degrees, strategic points of interaction, or secluded spaces for quiet conversations. These behavioural patterns and the general accessibility of buildings are governed by the application of access control. Access control systems introduce restrictions on the permissible movement degrees and help protect valuable parts of the building from unauthorised entry. However, it is often practicality and cost objectives, rather than security concerns, which guide the development and deployment of access control systems.

Access control systems generally require two tasks to govern accessibility: authentication and authorisation. Authentication verifies the identity of the person requesting access and authorisation determines whether access should be granted according to the specified access control policy. Access control must be performed at all entry points to protected facilities, which generally means that authentication must also be performed. Repetitive re-authentications require attention from the users and reduce the usability of the overall system. Moreover, the cost of installing authentication devices on every entry point may be prohibitive in many contexts. This combination of poor usability and cost generally results in a coarse granularity of the access control system.

Access control systems with low usability are at the risk of being compromised by the very users it is trying to protect. Users may leave doors open or share access cards or passwords to reduce the repetitive re-authentications. Mark Weiser states, in his vision of ubiquitous computing, that technology must be *calm* [2] in order to allow users to focus on their primary tasks. This implies that any authentication or authorisation technology should require minimal attention from the users. Further, the advances in technology have made us used to, and even come to expect, that technology is ever-present. Weiser predicted that: *"Next comes ubiquitous computing, or the age of calm technology, when technology recedes into the background of our lives."* [3], and technology has receded, not only into

the background of our lives, but into the very environment in which we interact. Technology embedded into the environment in this way makes the environment *smart* and may be defined as *"a small world where all kinds of smart devices are continuously working to make inhabitants' lives more comfortable"* [4, p. 3].

Smart devices embedded in the environment in this way may help in the authentication of inhabitants and could be equally suited for determining the location of these inhabitants. As such, indoor positioning systems have seen an increase in popularity in recent years. In particular, tracking of inhabitants in indoor environments have become vital in hospitals to locate and page staff, in homes for elderly people, and in industry for applications in logistics, warehousing and automation. Ideally, the provision of such services are completely transparent to the users, who simply observe that services are available as they are needed, for example, doors unlock and open when approached and lights that automatically turn on when a room is entered.

In this thesis we take a new approach to authentication, by combining traditional access control systems with the sensing technologies and tracking capabilities offered by smart environments. Our approach is called *Persistent Authentication for Location-based Services* and extends the underlying authentication model *PAISE*, proposed in earlier work by Kirschmeyer et al. [5]. The goal in persistent authentication is to enable the secure provision of location-based services through calm authentication of mobile users in a smart environment. Our objective is to shift the current authentication paradigm from a single discrete event to a continuous session. We accomplish this by utilising the contextual awareness provided by the smart environment to track principals from the point of initial authentication to the point where authorisation is requested by the location-based service.

1.1 CONTEXTUAL AWARENESS

Contextual awareness is generated by smart devices that sense their environment and record the aspects of their surrounding context. Liberman et al. [6] state that context depends on where the boundary between the system and the environment is drawn, as this affects what is considered explicit and implicit in the system. In this way, context can be any measurable and relevant information and each type

of context is a distinctive element. For smart environments, Schilit et al. [7] define the three most important aspects of context as being: where you are, who you are with, and what services are nearby. Moreover, Schilit et al. find the temporal changes in contexts are important as the same sensory data may have different meanings according to changes in both time and location.

The key application for contextual awareness in persistent authentication is to estimate the location of principals as they interact with services in the environment. We use the contextual awareness provided by the pervasive technologies to infer the location of principals in consecutive sensory measurements. This leads to the matching problem, which is the problem of identifying which principals are corresponding in two following measurements. This requires a similarity metric to indicate the level of correlation between the principals. In addition, to avoid the possibility of comparing the similarity metrics of all principals in the environment, the system must predict the location of each principal in the next measurement.

The similarity between two individuals can be described by their biometrics, that is, their intrinsic human traits. For humans, learning to recognise people just by looking at their face is a fundamental skill that we acquire early in childhood [8] [9] and which remains an integral part of our social interactions throughout life. We use faces to identify people in a group and in addition faces provide us with information about peoples demographic, including sex, race, and age, as well as their emotional state. We perceive this information effortlessly and apply it without conscious thought in our interactions with other people.

When asking a child to draw a face, the result might be two circles for the eyes, a curved line for the mouth, and perhaps a dot for the nose. While seemingly simple, these archetypical features represent the basic forms of a face and resemble the features used in computer vision face detection methods [10] [11]. With these methods, facial recognition can be used to provide a similarity metric, and have several advantages over other biometrics such as fingerprints, handprints and iris recognition, as it is non-intrusive and can be captured remotely, without user interaction.

In persistent authentication we use facial recognition as a similarity metric to solve the matching problem by periodically verifying the identity of tracked principals. To ensure a calm authentication process, the facial features are captured at a distance with no interaction required by the principals. The main challenge with

this approach is that the process is not completely reliable: the biometric system may fail to identify a principal, or conversely, may make an incorrect identification. The frequency of these errors depends on the discriminative power of the biometric and the quality of the acquisition process, which is affected by adverse environmental conditions, such as dust or poor luminosity. These factors are further compounded when using remote biometrics, as the quality and resolution of the captured biometric sample are significantly lower due to the uncontrolled acquisition process.

To increase the reliability of the biometric verification, multiple biometric characteristics can be used simultaneously. Common for such an approach is that the combined system is generally more robust than each of the individual components. Combining biometrics in this way is known as sensor fusion, and a common challenge in this field is that the results from evaluating the different biometric characteristic are usually incompatible, as they have different score ranges as well as different distributions. Consequently, a sensor fusion technique is needed.

In this thesis we present *error-rate-based fusion*, a novel fusion technique that transforms individual scores from different biometric systems into objective evidences and combine them using Bayesian inference. Error-rate-based fusion uses the false acceptance and false rejection rates of the biometrics systems to generate a confidence value that represents the probability that a principal has been correctly identified.

This combination of persistent authentication, remote biometrics and sensor fusion forms a robust situational protection measure capable of exploiting contextual information, provided by pervasive technology, to address the complex problems of protecting buildings against human threats and providing secure location-based services. In essence, the persistent authentication system makes it possible to perform informed decisions based on the user's current context, for instance, detecting that the cleaning personnel are accessing a restricted area, or that the carrier delivering goods does not go beyond the loading area.

1.2 EUROPEAN SEVENTH FRAMEWORK PROGRAMME

In 2010, the European Union funded the Resilient Infrastructure and Building Security (RIBS) project under the Seventh Framework Programme to support the specification of requirements for effective and affordable protection measures. Focusing on commercial buildings in an urban environment, the RIBS project was built on the idea that eliciting a set of requirements and making it available to technology developers would support the development of more satisfactory security products.

In order to achieve this aim, the RIBS consortium deployed the expertise required to accurately analyse the problem of requirements engineering, considering the various factors that could affect the quality of the results. Knowledge in the areas of business, security and crime science, laws and ethics, architecture and several fields of science and technology was leveraged to address it.

Over a period of thirty-six months, the RIBS consortium developed and adopted a multi-disciplinary approach to meet the following objectives:

- To develop a set of requirements for new counter-terrorism protection measures for commercial buildings.
- To develop a set of measurements that can be used to evaluate the level of protection offered by candidate security measures proposed to be implemented in buildings and infrastructures.

The tasks carried out to achieve these two objectives contributed to meeting a broader objective:

- To develop and apply a methodological framework that can support the development of requirements for physical security measures considering a range of constraints and objectives.

Several elements within the description of work limited the scope of the project to prioritise existing buildings and therefore retrofitting and to consider a range of threats, namely chemical, biological, explosives, and insider or intruder (CBE-I) agents. This thesis concerns the protection measures developed as part of the RIBS project to detect and deter insiders and intruder threats.

1.3 MOTIVATION

Most of the counter-terrorism technologies currently deployed in public spaces are concentrated at a few sites including airports, embassies and major event venues. In most cases, the implementation of security procedures has noticeably affected our behaviour, and simple items such as water bottles are now on the lists of prohibited objects. In addition, the security technologies acquired to protect us are relatively expensive, and their intrusiveness and poor efficiency undoubtedly more widely known than their effectiveness.

In comparison, few counter-terrorism measures exist outside these highly controlled places to protect the population. The large majority of commercial buildings and cultural venues, open to the public, are fitted with anti-theft measures, not counter-terrorism ones. Whilst intelligence agencies claim to have successfully disrupted a number of terrorist plots the fact remains that when criminals are not caught before reaching their destination, terrible damage occurs. Such is the case with the double bomb attack on the British Consulate and the HSBC headquarter in Istanbul, and the recent attack on the Marriot hotel in Islamabad.

The cause of the general unpreparedness can be found in the lack of incentives for most organisations to invest in counter-terrorism equipment. The frequency of serious terrorist attacks is relatively low in comparison with the number of potential targets, making security investments difficult to justify. At a time where many European governments reduce public expenditure, it seems justified to ask whether the cost of security is too high. Additionally, the cost of security is incurred in more than monetary expense, and the need for even more security systems in society is debatable.

Conversely, there is a need for effective integrated protection measures distributed in the facilities that are non-intrusive to the public and affordable to the facility owners. At the same time the measures must take into account the evolving resources and methods of modern terrorists. Motivated by these factors we present persistent authentication, an effective situational protection measure for smart environments that detects unauthorised entry and access, and provides early warning of attacks.

There is a growing interest in surveillance applications, due to the prevalence of cheap sensors and processors at reasonable costs. Further, the growing maturity of algorithms and processing techniques over the past few years enables the

application of remote surveillance of unattended environments. For sensor-based person tracking, the type, range and density of sensors embedded into the environment depends on the threat scenario and the characteristics of the protected area. Simple motion detection sensors may be appropriate in sparsely populated areas, while multiple sensors, possibly of different types, may be needed in more densely populated areas. Commonly, the most prevalent sensor type found in critical facilities today are closed-circuit television cameras.

Thus, to ensure the viability of the approach, the implementation must integrate with existing authentication and sensor technologies to minimise cost objectives and ensure that the privacy of monitored principals are not unnecessarily compromised.

In persistent authentication we utilise the authentication sessions acquired from the authentication system, the contextual awareness provided by the smart environment and the confidence values from the sensor fusion technique to detect anomalous behaviour and ensure the secure provision of location-based services. In this way persistent authentication can help protect against terrorism, and indirectly against other types of (organised) crimes. We recognise the difference between terrorism and volume crime; however our approach is sufficiently robust to allow its application to a wider range of security problems.

1.4 OBJECTIVES

The goal of this thesis is to develop a novel approach for person authentication in smart environments to improve the resilience and security of buildings in Europe. The proposed algorithms are implemented in close cooperation with an European financial Institute to ensure the viability of the approach. Case studies and evaluations are performed to test the system and assess the performance. The case studies form the basis for the evaluation of the systems persistence, robustness, and scalability, whereas the evaluation on publicly available datasets ensures the reproducibility of the results. The main objectives of this thesis are:

1. Shifting the current authentication paradigm from a single discrete event to a continuous session by developing a secure, calm and persistent authentication methodology that combines traditional authentication systems with the sensing capabilities offered by smart environments.

2. Providing contextual awareness to location-based services by ensuring that multiple mobile principals are efficiently and robustly authenticated and tracked while present in the smart environment.

3. Learning contextual and behavioural patterns of the occupants from multiple sources, including the movement and interaction behaviours of principals in the environment to detect diverging patterns.

1.5 CONTRIBUTIONS

This thesis presents four contributions: (1) development and evaluation of persistent authentication, (2) introduction of remote biometrics for continuous user authentication, (3) formulation of error-rate-based fusion of biometric systems, and (4) implementation of contextual and behavioural recognition techniques.

1. The contributions to persistent authentication constitute the implementation of a robust, non-invasive, authentication system for mobile users in a smart environment that uses closed-circuit television cameras. The result is a calm approach to authentication, where users are transparently authenticated towards the system.

2. Remote biometrics for continuous user authentication are introduced to solve the matching problem by periodically verifying the identity of tracked principals. The contributions include research into the biometric characteristics suited for continuous evaluation and the implementation and evaluation of these algorithms.

3. A novel sensor fusion scheme is presented to solve a common problem that occurs in sensor fusion, namely, that the evaluation of multiple biometric characteristics produce results that are incompatible, due to different score ranges and different probability distributions. The contributions constitute the formulation and evaluation of an error-rate-based fusion technique.

4. Data exploration techniques are introduced to identify contextual and behavioural patterns in the data. A dedicated tool is developed to process and structure very large quantities of data, to learn the prevalent patterns with

regards to the contextual awareness and the spatial configurations of the environment. The findings are integrated with state of the art data visualisation tools to present the information in a meaningful way.

1.6 THESIS OUTLINE

The presented research is organised in the following way. Part II provides an overview of the state of the art to frame the further research. Chapter 2 discusses authentication factors in smart environments, continuous authentication and biometric characteristics for use in biometric authentication. Following is a discussion of multi-factor authentication and remote biometrics and finally an overview of fusion techniques and the problems associated with normalisation and transformation of score level fusion. Chapter 3 concerns person tracking in smart environments and provides an overview of tracking techniques and a classification of different system topologies for positioning systems.

Part III presents the persistent authentication model, the definition of authentication and authorisation zones and the statistical relationship between the measurements and the environment. The pseudo-code for the persistent authentication algorithm is presented and methods for anomaly detection discussed. Chapter 5 presents error-rate-based fusion, the operational phases of the fusion strategy and how to combine the output of multiple biometric experts into objective evidences. In addition, the effect of quality on fusion is discussed.

Part IV details the algorithms and techniques used in the implementation of the camera-based persistent authentication prototype, the tracking algorithm and the implementation of remote biometric experts.

Part V discusses the evaluation of persistent authentication and consists of a combination of case studies, scenario-based validations and evaluation on public-domain datasets. The implementation of the prototype is evaluated on the basis of its persistence, robustness, and scalability.

Part VI presents our conclusions and contributions and outlines the directions for future work.

Part II

STATE OF THE ART

2

AUTHENTICATION IN SMART ENVIRONMENTS

Authentication is the process of verifying a user's identity to ensure the integrity, confidentiality, and availability of a system. The authentication process involves three phases: enrolment, presentation and evaluation. The enrolment phase precedes the other two phases and typically happens only once. In this phase information about the users are acquired and stored within the authentication system for later use. In the presentation phase, which occurs every time the user's identity needs to be verified, the user's authentication information is presented to the system. In the evaluation phase, the newly presented authentication information is compared to the enrolment record and a decision whether to authenticate the user is made. Traditionally, for a static authentication process, this decision is binary: accept or reject.

To implement calm authentication, users must be continuously authenticated in an unobtrusive way, i.e., not requiring their participation. Alternatively, users can be authenticated at a single, strategic, location and have the smart devices embedded in the environment facilitate their authentication by providing the authentication information to the location-based services as they are requested.

Existing authentication methodologies outline three basic factors that can be used to authenticate principals [12]:

1. Something the user *knows* (e.g., a password or passphrase).
2. Something the user *has* (e.g., an access card or smart-card).
3. Something the user *is* (e.g., biometric characteristics, such as a fingerprint).

These factors are considered device-centric, if the authentication is based on possession, i.e., something the user *has*, or user-centric if the authentication is based on knowledge or inherence, i.e., something the user *knows* or *is*. The factors may either require active participation, such as swiping a smart-card or using a fingerprint scanner, or no interaction, for instance carrying a wireless authentication device or being subject to a remote biometric scanner. A multi-factor

authentication system combines two or more of these factors and provides a higher degree of assurance than a single-factor system [13], as it is more difficult to forge or obtain several authentication samples at the same time. By definition true multi-factor authentication requires two authentication factors from different categories. Combining multiple factors from the same category also increases the security of the system, but it does not constitute multi-factor authentication [14].

Bardram et al. [15] [16] note that ubiquitous computing implies a shift in the use of computers, moving from a 'one-to-one' to a 'many-to-many' relationship between users and computers. This change in paradigms creates new usability challenges for computer security, especially user authentication, as the contemporary user authentication schemes are not calm. When using a personal computer for everyday tasks this is only a minor inconvenience, but in some work environments it poses substantial usability problems.

Bardram et al. base their research on studies of medical work at several large Danish hospitals [17] [18], where they observed that it was not uncommon that clinicians would log in 30 times a day on various devices. As a result, users would avoid logging out, share passwords or hand over sessions to one another, which compromised the security of the system. To make the authentication system calm, Bardram et al., proposed *location-based authentication*, a concept that builds upon the founding work of Denning et al. [19]. They define a proximity-based user authentication mechanism using wireless tokens that supplements the existing authentication methodologies with information about the location of users. This *"allows users to be authenticated on a device simply by approaching it physically"*. Consequently *location* can be integrated as the fourth factor in the classical user authentication methodologies.

2.1 CONTINUOUS AUTHENTICATION

Traditional authentication systems provide point-of-entry authentication, e.g., the username and password combination, where the user's identity is not verified subsequently after the initial authentication. In many situations, such discrete authentication mechanisms are not sufficient. Corner and Noble [20] [21] [22] investigate this problem and define traditional authentication mechanisms as *persistent* because they rarely limit the duration that the authentication is valid. As

a result, a user may leave a device logged in for a prolonged period of time and anyone who gains physical access to the device may usurp the authentication of the original user.

Consequently, Corner and Noble define a *continuous authentication* mechanism, where all data in the system are encrypted and a small authentication token, worn by the user, is required to provide access to the encrypted data. The token stores the cryptographic keys and continuously authenticate the user's presence over a short-range, wireless link. This combination of continuous authentication with the evaluation of the user's location ensures that access is only granted when the user and the token is in close proximity to the system. Conversely, Corner and Noble define this authentication approach as *transient* as the system can react to any changes that might affect the user's authentication session.

Typically, continuous authentication systems implement multi-factor authentication to increase the level of security and to lessen intrusiveness. To minimise the disruption of the user's work, Corner and Noble propose the use of a knowledge-based factor for point-of-entry authentication and an unobtrusive possession-based factor for all subsequent verifications. As a result, we can distinguish between the *active stage* of continuous authentication, where the initial authentication of the user occurs, and the *passive stage*, where the continuing presentation and evaluation of the user's identity are performed.

The definitions of continuous authentication by Corner and Noble and the concept of location-based authentication by Bardram et al., are device-centric as both require the presence of an authentication token. This puts additional restrictions on the users, e.g., that they have to wear the authentication token and creates problems when tokens are forgotten, borrowed, lost, or stolen. Furthermore, it shifts the problem of authentication from the user to the token, which introduces additional vulnerabilities in the system concerning the integrity of the device and its communication.

Wireless communications are susceptible to a range of attacks, including man-in-the-middle, proxy and replay attacks from adversaries intercepting data transmitted between devices. A detailed discussion of these vulnerabilities are beyond the scope of this thesis, but it is noted that the impact of these attacks are well studied in the field of computer security and that the severity of these vulnerabilities can be mitigated by implementing secure protocols, for instance based on distance bounding [23], verifiable multilateration [24] [25] or constrained chan-

nels [26]. Still, any form of possession-based authentication will provide a lower level of assurance compared to an user-centric authentication scheme. Even so, recent research [27] are exploring the viability of tokens implanted directly into the human body, which makes these tokens considerably harder to lose.

User-centric authentication may be based on knowledge-, inherence-, or location based factors. For a calm authentication process, the factors have to be presented and evaluated in an unobtrusive way, which, as noted by Bardram et al., makes knowledge-based factors ill suited for calm authentication, as presenting knowledge-based factors requires active participation from the user. Consequently, most approaches utilise inherence- and location-based factors. Facial recognition is one of the more common methods and can be used both in the active stage, as point-of-entry authentication, or in the passive stage as a remote biometric, i.e., sampled remotely at regular intervals and used for continuous authentication.

2.2 BIOMETRIC AUTHENTICATION

A biometric characteristic, such as a face or fingerprint, are the intrinsic traits of humans and allow direct verification of users instead of relying on devices that must be carried or passwords that must be remembered. The biometric characteristic, also called the modality, has the discriminative power to differentiate users in a large group. Fingerprints, palm prints, DNA, and iris patterns possess high discriminative power and are defined as hard biometrics, while hair colour, skin colour, gait, height, and weight have low discriminative power and are defined as soft biometrics. In addition, soft biometrics such as gait, are considered behavioural characteristics as they are related to the pattern of behaviour of a person.

Biometric authentication is performed by processing a sample of the biometric characteristic with a specialised algorithm, know as a biometric expert. The expert extracts a small amount of data containing the features of the characteristic, e.g., the ridges and spurs of a fingerprint, which is called the biometric feature. This feature represents the unique aspect of the modality and is used for comparison against a biometric reference, which is part of the biometric enrolment database. This database links the true identity of the person to previously

captured biometric samples for that person. The result is a *comparison score*, or *similarity score*, that reflects the expert's confidence in the identity of the person.

In biometric authentication the objective is to verify the claimed identity of a principal. Alternatively biometrics can be used to perform identification, i.e., comparing the acquired biometric sample with the entire enrolment database and returning the best match.

Tistarelli et al. [28] state that currently one of the most urgent research topics is in distributed sensor networks that transparently use biometrics that require no actions from the end-user. Further, the authors note that one of the main limitations of the precursors of today's biometrics are the need to keep the device in close contact with the subject to capture the biometric sample.

Klosterman and Ganger [29] define a *continuous biometric-enhanced authentication* mechanism, that uses biometric authentication based on facial recognition to periodically re-authenticate users who are logged on to the system. If, at some point, the biometrics of the user sitting in front of the monitor does not correspond to the biometrics of the authenticated user, then the session is revoked and a re-authentication required. In this way, facial recognition is used in a similar fashion as the cryptographic tokens proposed by Corner and Noble for device-centric continuous authentication.

Klosterman and Ganger note that biometrics differ from most knowledge- and possession-based factors in the way they are evaluated. Knowledge-based factors, such as passwords, are encrypted and compared on a byte-by-byte basis and possession-based factors rely on digital inter-device communication, which both returns a boolean result: true or false. In contrast, evaluation of biometric characteristics does not produce such clear results due to variations in the measured features and in the measurement environment, such as variation in facial expression, illumination or background for a facial recognition system. Rather than a boolean answer, then the similarity score resulting from a biometric evaluation is dependent on the similarity between the presented biometric sample and the samples captured during the enrolment phase.

A biometric system authenticates a principal by comparing the similarity score with the system's *operating threshold* . If the similarity score generated by the biometric expert meets this threshold then the principal is considered a genuine user and authenticated. Conversely, if the similarity score does not meet the threshold then the person is considered an impostor and the request is rejected.

In this process, two types of error can be committed: falsely rejecting a genuine user or falsely accepting an impostor. The associated error rates are called the False Rejection Rate (FRR) and the False Acceptance Rate (FAR) respectively, and are important measures to assess the system's performance.

Point-of-entry authentication may utilise intrusive biometrics, such as fingerprints, iris- or retina-scans, which have very high discriminative power, whereas continuous authentication requires remote biometrics, captured in an unobtrusive fashion. In the active stage of continuous authentication the setup and capturing environment can be controlled to ensure more homogeneous capture of biometric samples, this may include enforcing restrictions on the position, pose and expression of the user or to ensure that the illumination and background of the scene is constant. In contrast, in the passive stage of continuous authentication no restrictions are placed on the users or the environment and consequently the passive stage will have a higher level of false rejections and false acceptances [30]. Typical a multi-factor approach is used in the passive stage of continuous authentication to reduce the overall error-rate.

2.3 MULTI-FACTOR BIOMETRIC AUTHENTICATION

The main challenge in biometric authentication is that the process is not reliable: an expert may reject a genuine user, or conversely, an expert may accept an impostor. A biometric expert may have insufficient discriminative power, especially within a large group [31], or adverse environmental conditions, such as dust or poor luminosity, can affect the quality of biometric acquisition. These factors are further compounded when using remote biometrics as the quality and resolution of the biometric acquisition is significantly lower due to the uncontrolled acquisition process.

The reliability of biometric authentication can be improved by employing multiple biometric experts and combining their outputs. Common for such an approach is that the combined system is generally more robust than each of the individual components. According to the standard report ISO/IEC TR 24722 2007 [32] the combination can be any of the following types:

- Multimodal: using multiple different biometric modalities, e.g., combining capture of fingerprints and faces.

- Multi-sensor: using multiple sensors for capturing samples of one biometric instance, e.g., capturing faces in both the infrared spectrum and the visible light spectrum.
- Multi-presentation: using multiple presentation samples of one instance of a biometric characteristic, e.g., several frames from a video camera of a face image (possibly but not necessarily consecutive).
- Multi-instance: using multiple biometric instances within one biometric modality, e.g., capture of both the left and right iris.
- Multi-algorithmic: using multiple algorithms for processing the same biometric sample.

Sim et al. [33] present a continuous authentication system based on multimodal biometrics in a Bayesian framework. Their approach integrates results from a fingerprint reader and a facial recognition system. The authors combine the biometric modalities with temporal information which allows the system to evaluate the probability that a user is present even when there are no biometric observation available. Muncaster and Turk [34] explore a similar approach as Sim et al., but use a Dynamic Bayesian Network to achieve continuous authentication with multimodal and multi-presentation biometrics. The advantage of a dynamic Bayesian network is its ability to model more contextual information. Both Sim et al., and Muncaster and Turk focus on a controlled environment, such as a workstation, where an impostor hijacks a logged-in session.

Altinok and Turk [35] present an approach for temporal integration based on uncertainty propagation over time for a multimodal biometric system. Their method operates continuously by computing expected values as a function of time differences. The system generates continuous results in terms of confidence in the identity of the user, which makes it possible to adjust the security level accordingly, in real time. Experimental results with simulated data of face, voice, and fingerprints have shown that the system can provide continuous authentication results, which are consistently better than the individual components of the system. The authors conclude that comparing these preliminary results to a true multimodal database is very important for continued work in the field.

Niinuma and Park [36] propose a framework for continuous authentication that uses soft biometrics, that is biometrics characteristics that have low discriminative power, such as skin colour, height and weight. The proposed framework automatically registers soft biometrics in the active stage of authentication, i.e., when the user login, and fuse the soft biometrics with conventional authentication schemes, namely password and face biometric. The experimental results show that the proposed scheme has a high tolerance of the user's posture and the authors conjecture that the proposed method provides effective and robust continuous authentication. However, the authors make a number of assumptions about the pose of the users and the location of the body for appearance analysis, furthermore, occlusions are handled on a very ad hoc basis.

The acquisition of biometric samples is subject to variations caused by the changing conditions of the measurement environment and the evolving nature of the human traits. Thus, biometric authentication is not completely reliable and an expert may reject a genuine user or accept an impostor. The performance of a biometric expert is reflected in the rate of which the experts falsely rejects and accepts principals. Multi-factor biometrics increases the reliability of the authentication process and biometric authentication system that uses characteristics with high discriminative power are considered on par with traditional authentication factors.

2.4 FUSION OF BIOMETRIC EXPERTS

Combining the output of multiple biometrics is called fusion, and the methods used are known as fusion techniques. We distinguish between two broad categories of fusion techniques: *feature level* and *score level* fusion [37]. Feature level fusion combines biometric samples before an expert computes a similarity score and score level fusion takes place after the comparison stage. It is also possible to combine scores at an abstract level, such as at the rank or decision level [38], however, there is a risk of loss of information in such abstractions.

Although there is more information at the feature level, properly exploiting this information in fusion may result in the curse of dimensionality, i.e., as the dimensionality increases, the volume of the space increases so fast that the available data becomes sparse, and thus the problem becomes very difficult to solve.

Furthermore, not all biometrics are compatible at feature level as they do not have the same dimension, type or sampling rate, e.g., gait and fingerprint, or iris and hair colour. In addition, if the relationship between the feature spaces of the different biometrics is unknown, then the additional information available at feature level cannot be properly exploited and feature level fusion will not provide any advantage over score level fusion.

Score level fusion combines scores from multiple experts into a single score, upon which the verification decision is made. Score level fusion reduces the problem of complexity and allows different biometric experts to be used independently of each other. Consequently, score level fusion is generally the preferred technique as it provides a good trade-off in terms of complexity and information. Further, one can freely change sensors, algorithms, and data representations in the implementation of individual experts.

A basic problem exists in score level fusion: the scores of individual experts have different score ranges as well as different probability distributions, e.g., a fingerprint expert may return scores in the range $[0, 1000]$, and a facial recognition expert might return scores in the range $[-1, 1]$. Many researchers address this problem by normalising the score ranges and then combine the scores with a scalar function, such as the product or sum of scores. In other approaches, probability density functions of scores are estimated, and the results combined as product of likelihood ratios. Machine learning algorithms can also be trained on the concatenation of similarity scores to compute the combined score.

Nandakumar et al. [39], state that score level fusion can be divided into the following three categories:

> *Transformation-based score fusion:* The match scores are first normalised (transformed) to a common domain and then combined. The choice of normalisation scheme and combination weights is data-dependent and requires empirical evaluation.
>
> *Classifier-based score fusion:* Scores from multiple matchers are treated as a feature vector and a classifier is constructed to discriminate genuine and impostor scores.
>
> *Density-based score fusion:* Scores are converted to likelihood ratios which requires explicit estimation of the genuine and impostor score densities. The advantage is that density-based score fusion directly

achieves optimal performance at any desired operating point (FAR), provided the score densities are estimated accurately.

Kittler et al. [40] [41] developed a theoretical framework of simple transformation-based score fusion strategies, namely the sum, product, minimum, maximum, median, and majority vote rules. They show that these simple strategies are obtained from their framework under different assumptions, e.g., the sum rule is obtained if one assumes that experts are independent and a-posteriori confidence in a claimed identity is close to the a-priori confidence. They show that, among these simple strategies, the sum rule is the most resilient to estimation errors if Gaussian distributions of errors are assumed. Later, Kittler and Alkoot [42] show that, for heavy tail distributions of errors, the majority vote rule may outperform the sum rule.

Due to their attractive properties these simple fusion strategies are the foundation of many of the more advanced fusion techniques. In addition, they are commonly used as benchmarks for comparative purposes.

Kryszczuk et al. [43] combine voice and face recognition using a fixed rule, in which the quality of the biometrics are used when making an authentication decision. Fierrez-Aguilar et al. [44] propose a fusion method using support vector machine (SVM) based classifiers. In their approach, the fusion method essentially corresponds to the sum rule. Nandakumar et al. [45] propose a quality based fusion method, in which the fusion is implemented as the product of likelihood ratios. Similarly, Maurer and Baker [46] propose a fusion framework based on Bayesian networks. Fusion in this framework is proposed by multiplication of likelihood ratios from different experts.

Toh et al. [47] [48] propose the use of a multivariate polynomial as a fusion strategy to linearly combine experts scores. The training phase consists of progressively adding individual polynomial terms and testing whether the new terms reduce the error in the fusion process. This model can be considered as a generalisation of product, sum, and power rules. As a complementary work, Toh and Yau [49] propose the use of hyperbolic functions for score normalisation in the multivariate polynomial. The multivariate polynomial model is used in later work, where Toh [50], and Toh et al. [51] explore the total error counting rate for supervised classifier learning. By utilising a smooth functional approximation to the error counting objective, they are able to formulate a single-step solution,

which in empirical evaluations show promising potential in terms of decision accuracy and computing efficiency.

Density-based score fusion techniques, as the approach used by Toh et al., have interesting properties, as the performance of a biometric experts usually are expressed by their error rates. Thus fusion of experts can be based on combining error rates, rather than normalising scores.

Hanmandlu et al. [52] presents a multimodal biometric system based on error level fusion. They present two error level fusion strategies, one utilising the Choquet integral and another utilising t-norms. Their strategy is more similar to decision level fusion as the error rates are derived from the decisions made on individual modalities. However, their formulation of the Choquet integral mitigates several of the drawbacks associated with decision level fusion. The non-additive aspect of the Choquet integral is capitalised in the interaction between the error rates of two modalities at a time and additionally helps decide the order in which the decisions should be combined to yield better accuracy.

Li et al. [53] propose a fusion function, which is a ratio of FRR to the error rate (i.e., FRR and FAR combined). They transformed the problem of estimating the distribution of scores to the distribution of FAR and FRR. The choice of the fusion function is somewhat ad hoc.

The error rates of a biometric expert are closely linked to the quality of the biometric acquisition. Especially for remote biometrics, the quality of the captured samples may strongly affect the resulting similarity scores. Thus incorporating quality in the fusion technique is an important step to ensure robust fusion.

Bengio et al. [54] suggest a classifier-based score fusion technique using machine learning algorithms for quality dependent biometric fusion. In their approach a machine learning algorithm is directly trained on a training dataset, which consists of matching scores from all experts as well as the respective quality measures. They suggest a number of methods to estimate the quality measures from the training data. Poh and Bengio [55] propose quality measures based on the absolute difference between false acceptance rates (FAR) and false rejection rates (FRR) distributions (i.e., FAR and FRR computed as the functions of similarity scores), however, computing the difference this way causes a loss of the information present in the values of FAR and FRR. In follow-up work Poh and Bengio [56] address this, and propose to also include client dependent information in the fused score.

Chatzis et al. [57] describe several decision-level supervisors, which also take into account the quality of expert decisions. The proposed supervisors are constructed using clustering algorithms. A clustering algorithm attempts to divide a dataset into subsets, which are called clusters. In the author's approach, these clusters are considered as fuzzy sets, where fuzziness represents the quality of an expert decision.

Bigün et al. [58] [59] propose a fusing strategy, which is based on Bigün's earlier work [60], and have roots in risk analysis. They designed, what they call, an expert conciliation system, which takes into account the accuracy and the quality of biometric signals. The authors also propose adaptive fusion, where adaptivity is a function of quality of input signals. In follow-up work, fierrez-Aguilar et al. [44] and Fronthaler et al. [61] subject the proposed method to a wide range of experimental data to test the performance and viability of the approach.

Poh and Kittler [62] propose an unified Bayesian framework for expert fusion, which incorporates quality measures. This framework encompasses many previously proposed quality-based fusion approaches. The unified model consists of two sets of Bayesian networks. One of the sets represents generative experts while the other represents discriminative experts. A generative expert first computes the similarity scores and then computes the class using Bayesian inference. A discriminative expert directly computes the class, and such an expert is usually based on a machine learning algorithm. Both generative and discriminative experts are further classified as non-quality, feature-based quality, and cluster-based quality experts. In all types of experts, fusion among multiple experts is proposed using the sum of normalised log likelihood ratio scores.

Rodrigues et al. [63] explore the security of multimodal biometric systems when one experts is spoofed. Their experiments show that for traditional fusion schemes, i.e., log likelihood ratio (LLR) or sum rule, a forger can crack a multimodal system by spoofing only one of the biometrics. This has implications for all multi-factor systems, however, the authors found that hardening the fusion process creates a trade-off between security and accuracy. Spoofed samples can be assumed to be of poorer quality than genuine samples, thus considering quality in the fusion strategy may mitigate some of the risk.

We note that most of the existing solutions for score level fusion are either based on ad hoc assumption or not well-understood in terms of fusion optimality.

For instance, a density-based fusion approach is typically based on a parametric model that serves as an analytical tool for converting raw scores to the probability of being genuine (or impostor). Similarly, it is difficult to prove that a machine learning algorithm is the optimum strategy to combine scores. Furthermore, the choice of normalisation scheme and combination of weights is data-dependent and requires extensive empirical evaluation.

2.5 SUMMARY

In authentication the claimed identity of a user is verified to ensure the integrity, confidentiality, and availability of a system. Existing authentication methodologies outline three basic authentication factors: something the user knows, (e.g., a password or passphrase), something the user has (e.g., an access card or smartcard) and something the user is (e.g., biometric characteristics, such as a fingerprint). In smart environments an additional fourth authentication factor is considered, namely the location of the user.

Traditional authentication systems are considered persistent as they rarely limit the validity of the authentication session. Conversely, sensor-based authentication systems are considered transient as they allow continuous authentication of the users. Continuous authentication systems typically implement multi-factor authentication to increase the level of security and to lessen intrusiveness. For a calm authentication process, these factors have to be presented and evaluated in an unobtrusive way, which lends itself to the use of user-centric authentication factors such as knowledge-, inherence-, or location-based factors.

Biometric characteristic, such as a face or fingerprint, are the intrinsic traits of humans and allow direct verification of users. Fingerprints, palm prints, DNA, and iris patterns possess high discriminative power and are defined as hard biometrics, while hair colour, skin colour, gait, height, and weight have low discriminative power and are defined as soft biometrics. The reliability of biometric authentication can be improved by employing multiple biometric experts and combining their output.

Combining the output of multiple biometrics is called fusion, and the methods used are known as fusion techniques. The main challenge in biometric fusion is that different experts generate matching scores in different domains, and these do-

mains usually follow different probability distributions. Therefore, score normalisation and transformation are required to make the scores compatible, which are error prone processes. Moreover, the existing parametric models assume a certain distribution of scores, such as a normal distribution, which also introduces errors in the fusion process.

3

TRACKING IN SMART ENVIRONMENTS

In its simplest form, tracking can be defined as the problem of estimating the trajectory of a principal moving through a smart environment. The trajectory is estimated by finding the *state* of the principal in each sensor measurement. The state represents the principal by a number of parameters such as position and, depending on the tracker domain, acceleration, velocity, orientation, and appearance.

The first step in tracking is to initialise the tracker, e.g., an appropriate model containing the position and other contextual information of the principal, must be established. Next, the motion of the principal is tracked. This implies a way of establishing the new locations of the principal and determining the correspondences between these in consecutive sensor measurements. Finally, as any tracker eventually fails, detecting the occurrence of these failures and reacting accordingly is important for the robustness of the tracker. Consequently, a tracker consists of the following:

Model. The tracking algorithm represents the state of the target principal by a model. This model can be either static, thus remaining fixed during tracking, or it can be adaptive, meaning it accepts new information and updates during tracking.

Motion estimation. Given the state of the principal in previous frames and the model, tracking estimates the motion of the principal by fitting the model to the current measurement using some estimation algorithm. Generally, only the vicinity of the previous state is explored in this step.

Failure. Tracking failure is a sudden incorrect estimation of the principal's state. Tracking failures are typically caused by poor sensor measurements, due to environmental factors, such as reflections, occlusions or appearance changes. Furthermore, the principal may move outside the range of the sensors, thus terminating the tracking.

In state-of-the-art trackers, the model is often adaptive and thus updated with new information during tracking. This allow the trackers to handle changes in the state of the principals or in the environment. The drawback of adaptation is drift, i.e., that errors in the update may accumulate over time, reducing the correspondence between the target and the model.

Thus, formally put, trackers characterise the target principal in every sensor measurement by several parameters (e.g. position, orientation, appearance), which forms the state of the principal. A temporal sequence of such states defines the trajectory of the principal and the difference between two consecutive states defines the motion of the principal [64]. An illustration showing the trajectory of principal a (solid line) and the estimated trajectory (dotted line) derived from a number of noisy sensor measurements $(t_0, \ldots, t_7)$ is seen in Figure 1.

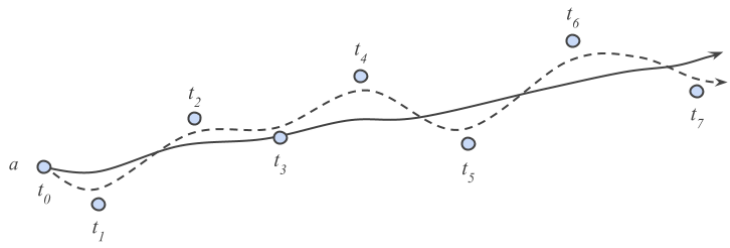


Figure 1.: Illustration of the actual and estimated trajectory of principal a.

To determine the state of the principals, smart environments prevalently employ either a device-centric approach based on wireless devices [65] or a user-centric approach using vision-based devices [66]. Wireless devices rely on hardware level communication between signal transmitters and measuring units, with the transmitters either placed in the environment or carried by the users. The context of the users are determined based on signal measurements such as time of flight, angle, and signal strength. Vision-based devices are a diverse field and include closed-circuit television cameras (CCTV), stereo-vision cameras, infrared time-of-flight (TOF) cameras and lasers. Vision-based systems rely on segmentation and feature extraction techniques to determine the context of the users. In addition, several less common, but novel, approaches can be used to determine the state of the principals in smart environments, namely: passive infrared mo-

tion detectors; acoustic detectors similar to sonars or based on triangulation [67], and pressure sensitive floors [68].

3.1 TRACKING WITH WIRELESS-BASED DEVICES

Wireless tracking systems consists of at least two separate hardware components: a signal transmitter and a measuring unit. Communication involves the transmission and reception of signals between these components.

In early work on wireless devices Drane et al. [69] state that positioning systems may be classified based on either: the location positioning algorithm, i.e., the method of determining location of users using various signal measurements, such as time of flight, angle, and signal strength; or the sensor infrastructure, i.e., the wireless technology used to communicate with the devices in the environment or carried by the users.

Based on this classification Drane et al., define three different system topologies for positioning systems:

1. *Remote positioning*, where receivers at one or more static locations measure a signal originating from, or reflecting off, the object to be positioned. These measurements are communicated to a central site where they are combined to give an estimate of the position of the object.

2. *Self-positioning*, where the positioning receiver makes the appropriate signal measurements from geographically distributed transmitters and uses these measurements to determine its position.

3. *Indirect positioning*, where a data link is used to send position measurements from the self-positioning receiver to a remote site or vice versa. A self-positioning system that sends position data to a remote location is referred to as *indirect remote positioning*, and a remote positioning system transmitting an object's position to the object is referred to as *indirect self-positioning*.

Fang [70] describes a remote positioning system that relies on the time of arrival (TOA) of transmitted signals for positioning. The distance from a mobile target to the measuring units is directly proportional to the propagation time,

thus with at least three reference points it is possible to compute the intersection points of circles formed by the time of arrival measurement. This computation is solved either with a geometric method or by minimising the sum of squares of a non-linear cost function.

Torrieri [71] propose a method to determine the relative position of a mobile transmitter by measuring the time difference on arrival (TDOA) at which the signal arrives at multiple measuring units. The TDOA measurements at two measuring units are combined to restrict the possible transmitter location to a hyperboloid with the two units as foci. The transmitter location is estimated from the intersections of three or more independently generated hyperboloids determined from at least four measuring units.

Fang and Torrieri's methods describe some of the early work in wireless positioning, which were developed to determine the location of cellphone users. Fang and Torrieri assume line of sight between the signal transmitter and the receiver, otherwise they suffer from the multipath effect, which decreases the accuracy of the estimated location. As line of sight is difficult to guarantee in an indoor positioning system, Ng et al. [72] instead propose to measure the attenuation of the emitted signal strength. Location estimation is made possible by using multiple measurements at several base stations, with the position of the stations as reference points. The accuracy of this method can be improved by utilising the pre-measured received signal strength (RSS) contours centred at the receiver, to mitigate the multipath effect [73].

Time of arrival, time difference on arrival and received signal strength form the basis of wireless positioning systems. The wireless technology employed in the systems are interchangeable, and can for instance be based on infrared light [74] or radio frequency, such as WLANs [75], RfiD tags [76] [77] or Bluetooth [78].

Delafontaine et al. [79] consider the application of Bluetooth technology as a tracking system. Their setup consists of a number of Bluetooth access nodes, installed at fixed strategic locations throughout the area of interest. Each node continuously searches for nearby devices. When a Bluetooth-enabled device is found, its MAC address is logged, such that the presence of devices at nodes can be recorded. From these records, the trajectory of an individual may be approximated as the spatiotemporal sequence of node observations of the device.

Fry et al. [80] propose a system, MASCAL, for enhancing the management of resources at hospitals during mass casualty situations. The system contains

three components: 802.11b RfiD tags, fixed transceivers that measure ambient 802.11b signal strength and a central geolocation server that computes location. The transceivers periodically measure the ambient signal strength between themselves and any other devices detected in the area. The geolocation server then calculates a reference signal strength topology for the coverage area, allowing the system to track patients, equipment and staff.

Device-centric tracking suffers from the same problems and limitations as device-centric authentication; it puts additional restrictions on the users and creates problems when the devices are forgotten, lost, or stolen. However, the prevalence of personal smart devices means that many users are willingly carrying wireless devices on their person at all times. This creates a number of interesting applications for device-centric tracking, especially with regards to classifying users movement patterns and behaviours.

3.2 TRACKING WITH VISION-BASED DEVICES

Vision-based tracking is the process of following an image element in a sequence of sensor readings. This may be accomplished by continuous, frame-by-frame, analysis of the element, by estimating the motion with respect to the target by tracking points or features on the target itself, or by analysing the change in every pixel from frame to frame.

Trucco et al. [81] formulate this process in terms of the *motion problem*, that is predicting the location of a tracked principal in the next frame, and the *matching problem*, that is re-identifying the principal within a designated search region. The motion and matching problems are the foundation of the *predict-match-update* framework, which forms the basis of many vision-based tracking algorithms.

In the predict-match-update framework, the state of the target principal in the current frame is used to *predict* the location in the next frame. The prediction returns a search region, also known as a region of interest (ROI), within which the principal will likely be. The search region is analysed and a *match* between the target principal and the region is found. In multi-target tracking, this match score is used to describe which principal best match each of the respective search regions. In essence, this is a data association problem where a similarity metric is

used to assign the respective trajectory to each of the tracked principals. Finally, as noise will almost always factor in the sensor measurements, an *update* of the current state combined with the previous states are performed to increase the robustness of the tracker. Figure 2 shows a model of the predict-match-update framework.

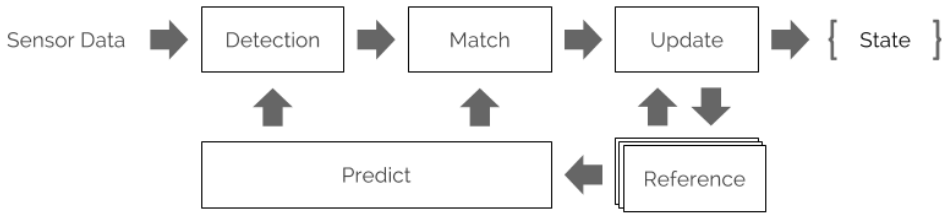


Figure 2.: Predict-match-update tracking framework

Trucco et al. formulate a set of design requirements for vision-based tracking applications such as the *predict-match-update* framework. These requirements are:

> *Robustness to clutter*: the tracker should not be distracted by image elements resembling the target being tracked.
>
> *False positives/negatives*: only valid targets should be classified as such, and any other image element ignored (in practice, the number of false alarms should be as small as possible).
>
> *Agility*: the tracker should follow targets moving with significant speed and acceleration.
>
> *Robustness to occlusion*: tracking should not be definitely lost because of temporary target occlusion (drop-out), but resumed correctly when the target reappears (drop-in).
>
> *Stability:* the lock and accuracy should be maintained indefinitely over time.

Consequently, a tracker needs to consider that a tracked principal may appear in cluttered environments and may be surrounded by other principals. The tracker needs to accommodate for this, and should not get distracted by background

clutter. In addition, the tracked principal may get occluded or disappear from the camera view for an arbitrarily long time and then reappear at any time and at any location. Finally, to be useful in interactive environments, the tracker needs to work in real time, at full frame rate. Therefore, the tracker must be extremely efficient.

The first three design requirements require robust detection of the principals in the first step of the predict-match-update framework. This is a classical computer vision problem and we give an overview of some of the different approaches in subsection 3.2.1. Next, for occlusions a common approach is to try and reduce the occurrence of occlusions by appropriate selection of camera positions, e.g., with ceiling mounted cameras most occlusions can be eliminated. Even though, occlusions cannot be completely avoided. One approach for dealing with the remaining occlusions is to leverage motion consistency, i.e., the trajectory and velocity of principals rarely change drastically from one frame to the next. Thus if an occlusion is detected, the tracker can predict the location of the principals after the occlusion. In subsection 3.2.2 we give an overview of some of the common approaches for mitigating the effect of occlusions through motion estimation. Finally, in subsection 3.2.3 we present notable related work with regards to vision-based tracking.

3.2.1 *Image Segmentation*

Image segmentation is the process of identifying and delineating objects in an image or video that share certain characteristics. Segmentation is based on observed data and relies on information present in each pixel in the image. This typically include the pixel values, the (x,y) position of the pixel and other information characterising local texture. The result of image segmentation is a set of segments that collectively cover the entire image, where each segment are similar with respect to some characteristic, for instance colour, intensity, or texture.

In person tracking, background subtraction is typically used as a method to segment moving regions in image sequences taken from a static camera, by comparing each new frame to a model of the scene background. The segmentation is performed by assigning one label to the principals and another label to their

surroundings. The principals are then referred to as the image *foreground* and their surroundings as the image *background*.

As an example, consider a binary label $w_i \in \{0,1\}$ that is assigned to each pixel $\mathbf{x}_i$ in an image, indicating whether it is part of a known background ($w = 0$) or if it belongs to the foreground ($w = 1$), determined by the recent history of each pixel $\mathbf{x}_1, ..., \mathbf{x}_n$. A typical result of such a foreground/background segmentation is shown in Figure 3. The top image shows a complex scene, captured by a surveillance camera, containing five principals annotated with circles. The bottom image shows the output, a black-and-white ($w = 0$ and $w = 1$ respectively) binary image. The white pixels in the binary image, known as the blobs, indicate the presence of a principal. The figure shows that all five principals have been correctly identified.

The quality of the labelling process is of great importance for the accuracy of the tracker, thus special care must be taken in assigning the labels. The first step in this process is to select an appropriate method for background modelling that fits the chosen application. Many background modelling methods have been proposed, each with their own strengths, weaknesses and intended applications. In a recent survey Bouwman [82] evaluates the notable advances in image segmentation and introduces a classification based on the following categories:

> *Basic Background Modelling:* in which the background is modelled using the average [83] or the median [84] or histogram analysis over time [85].
>
> *Statistical Background Modelling:* where the background is modelled using a single Gaussian [86] or a mixture of Gaussians [87] [88] or a Kernel Density Estimation [89]. Statistical models are used to classify the pixels as foreground or background.
>
> *Fuzzy Background Modelling:* which models the background using a fuzzy running average [90] or Type–2 fuzzy mixture of Gaussians, where the foreground detection is made using the Choquet integral [91] [92].
>
> *Background Clustering:* which assumes that each pixel in the frame can be represented temporally by clusters. Incoming pixels are matched against the corresponding cluster group and are classified according

Figure 3.: Background segmentation. For each pixel in the image a label w is inferred denoting the absence or presence of a foreground object.

to whether the matching cluster is considered part of the background. The clustering approach can be based on the K-mean algorithm [93].

Wavelet Background Modelling: which defines the background model in the temporal domain, utilising the coefficients of discrete wavelet transform (DWT) [94].

Background Estimation: where the background is estimated using a filter, such as the Kalman filter [95]. Any pixel of the current image that deviates significantly from its predicted value is declared foreground.

In the context of image segmentation Bouwman notes that these modelling approaches must consider the following issues: background initialisation, background maintenance, foreground detection, choice of the feature type (colour features, motion features, edge features or texture features) and the feature size (pixel, block or cluster).

When selecting an image segmentation method these issues determine the robustness of the method with regards to the critical situations faced in tracking using vision-based devices. The performance of the segmentation process deteriorates with dynamic changes to the background or the illumination of the scene, the appearance and pose of the principals, occlusions and obstructions in the field of view of the camera, and a number of other factors that affect the image acquisition. We identify four types of factors that affect performance: (1) technology, (2) environment, (3) user, and (4) user-system interaction, summarised in Table 1.

Image quality, compression algorithms and the variations between images acquired on different devices or the slight variations arising from multiple acquisitions on the same device all affect how the objects appear in the image. Similarly, changes in the illumination due to variations in natural or artificial lighting, and the presence of shadows and reflections makes similar objects appear less homogeneous, and, as a result, makes the task of correctly delineating these objects more difficult.

In addition, for tracking purposes, the movement of users, their appearance and pose, all affect the performance of the segmentation process. Additionally, occlusions, caused by interposing users or objects, makes delineating objects very challenging and may require additional information such as the depth of the

Table 1.: Factors affecting image segmentation

Type	Factor
Technology	Image quality, compression, heterogeneous acquisition.
Environment	Illumination (indoor, outdoor), shadows, reflections.
User	Movement, pose, appearance, cooperation.
User-System	Occlusions, alignment between camera axis and user.

image. This effect is further compounded as users, either on purpose or unknowingly, may create situations that maximises the impact of these factors.

One of the most prevalent methods for image segmentation is the Gaussian Mixture Model. This approach was originally proposed by Friedman et al. [96] and later Stauffer et al. [87] [88] extended this work by introducing efficient update equations.

A Gaussian Mixture Model is a pixel-level image segmentation method, in which the background model has a probability density function for all pixels in the image. Pixels from a new image is considered part of the background if their values are well described by the corresponding density functions. This approach requires appropriate values for the variances of the pixel intensity levels, since the variances can vary from pixel to pixel. In addition the pixel values often have complex distributions.

The Gaussian Mixture Model states that the probability that a new pixel $\mathbf{x}$ belongs to the foreground ($w = 1$) is given by:

$$Pr(\mathbf{x}|w = 1) = \sum_{k=1}^{M} \lambda_k \mathcal{N}(\mu_k, \Sigma_k) \tag{1}$$

where M is the number of Gaussians, $\mu_{1...M}$ and $\Sigma_{1...M}$ are the means and covariances of the normal distributions and $\lambda_{1...M}$ are positive valued weights that sum to one.

The combination of these normal distributions allows the Gaussian Mixture Model to describe multimodal probability densities. The multimodality of the

model means that complex backgrounds can be modelled and that the approach is robust to noise and gradual changes in environmental factors, such as illumination. However, the model present some disadvantages. Firstly, the number of Gaussians must be predetermined, which is a problem as using too few Gaussians means the method cannot accurately model backgrounds with fast variations. Conversely, using too many Gaussians unnecessarily increases memory and computation time. Secondly, the model requires good initialisations and is dependent on a series of training frames absent of moving objects.

To alleviate these limitations, numerous improvements have been proposed. These include intrinsic and extrinsic improvements to the model, dynamic selection of the optimal number of Gaussians, enhancements to the foreground detection and significant reduction in computation time and memory consumption. In the survey by Bouwman [82] the author presents an overview of these improvements to the original Gaussian Mixture Model method and evaluates the impact of the different approaches. The conclusion is that with the improvements the Gaussian Mixture Model presents a robust and accurate approach, that is shown to perform reliably in a number of different settings, and provides a good balance between performance and complexity compared to other image segmentation techniques. These findings are also reflected in work by Brutzer et al. [97] in their evaluation of background subtraction techniques for video surveillance and in work by Parks et al. [98] in their investigation and evaluation of the impact of post-processing on background modelling techniques. As a result, the Gaussian Mixture Model is widely used and is particularly suited for indoor surveillance applications where changes in the background and illumination occurs gradually.

Cucchiara et al. [99] propose a computationally lighter algorithm, that uses a running average to model the background. This model bases the value of each pixel in the background model on the recent history of images, either the previous n frames or a weighted average with the most recent frames having a higher weight. In essence, the background model is computed as a chronological average of each pixel's history. In their evaluation of background modelling methods Parks et al. [98] show that such simple modelling techniques can perform nearly as well as the more complex methods under the right conditions.

Oliver et al. [100] propose to use an eigenspace representation of the background. A principal component analysis is used on a sequence of n images to

compute the eigenbackgrounds. New objects are detected by comparing the input image with an image reconstructed via the eigenspace.

Elgammal et al. [89] propose the Kernel Density Estimation (KDE) method to deal with dynamic backgrounds like waving trees and rippling water by estimating the probability density function for each pixel using the kernel estimator, K, for the n, recent samples of intensity values, $\mathbf{x}_1, ..., \mathbf{x}_n$, taken consecutively in a time size window W. For foreground detection, the parameters of the model must be updated. Elgammal et al. propose to use two background models: a short term model and a long term model. The short term model adapts quickly to allow very sensitive detection and consists of the most recent n background sample values. The long term model captures a more stable representation of the scene background consisting of n sample pixels taken from a much larger time size window and thus adapts to changes slowly. Foreground detection is obtained by taking the intersection of the outputs of the short term model and the long term model as this eliminates false positives that are described in either model. Bouwman [82] note that the KDE method is more adapted for outdoor scenes where dynamic backgrounds appear but less suited for illumination changes.

3.2.2 *Motion estimation*

Knowing the current position of a principal and being able to anticipate that principal's trajectory is an important step for person tracking in highly populated areas. Trackers generally perform well when it is possible to observe the tracked principal in each time step. However, in real world environments, the trajectories of tracked principals often intertwine and the principals may become partially occluded. Long periods of occlusion can cause a tracker to lose a principal, or worse, the tracker may falsely identify the principal as a different person. In many situations, this problem can be avoided by using a motion model that gives an estimation of where the principal might be after the occlusion.

In vision-based tracking, motion may be determined using optical flow analysis [101] [102] [103] [104] [105] which calculates the motion vectors of pixels in the image. The algorithms assume that when a pixel moves from one frame to another, its intensity or colour does not change. This is a combination of a number of assumption about the reflectance properties and illumination of the scene

and is known as the *brightness constancy*. Defining *I(x,y,t)* as the intensity of a pixel *(x,y)* at a given time *t*, then the brightness constancy with flow *u(x, y, t), v(x, y, t)* can be written as:

$$I(x,y,t) = I(x+u, y+v, t+1) \tag{2}$$

This equation can be linearised by applying a first-order Taylor expansion to the right-hand side yielding the *optical flow constraint*:

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \tag{3}$$

In dense algorithms, solving the optical flow constraint yields the magnitude and direction of motion for every pixel in the image, whereas for sparse methods, only a region of the image is investigated.

When principals are occluded it is not possible to accurately calculate their motion, thus it is necessary to rely on a motion model to predict their position. One such model is the Kalman filter [95] [106], a well-known approach from control theory, which provides an optimal, recursive estimator of the state of a dynamic system.

The Kalman filter model assumes that the state vector $\mathbf{x}_t$, which contains the terms of interest for the system, at time *t* is evolved from the previous state *(t–1)* according to the equation:

$$\mathbf{x}_t = \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{B}_t \mathbf{u}_t + \mathbf{w}_t \tag{4}$$

With a measurement $\mathbf{z}_t$ given by:

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t \tag{5}$$

Where $\mathbf{F}_t$ is the state transition matrix that relates the current time step to the previous one and $\mathbf{H}_t$ is the measurement model that relates the state to the measurement. $\mathbf{u}_t$ is the control input vector and $\mathbf{B}_t$ is the control input matrix which applies the effect of each control input parameter on the state vector. $\mathbf{w}_t$ and $\mathbf{v}_t$ represents the process and measurement noise and are assumed to be independent normal probability distributions.

Kalman filters are widely used in computer vision for video tracking [107] and have been adapted [108] [109] to accommodate a variety of uses. However,

a basic problem with the Kalman filter remains, namely, that it is based on Gaussian densities which, being unimodal, cannot represent simultaneous alternative hypotheses. The observation models and target distributions can be highly non-linear and non-Gaussian, and, in addition, occlusions and complex interactions between users create ambiguities and overlap. To overcome these difficulties, the non-linear state dynamics can be solved with analytical linearisation [110], which is known as the Extended Kalman filter (EKF), or with statistical transformations [111], known as the Unscented Kalman filter (UKF).

The differences between the performance of the EKF and UKF are shown to be small [112] [113], and a thorough sensitivity analysis of both filters for the problem of GPS/INS attitude estimation revealed that the two filters give similar results even as a function of various design parameters such as noise, covariance tuning, sampling rate, and initialisation error [114]. Basically, the differences between the EKF and UKF become more significant as the non-linearity in the system increases [115].

Isard et al. [116] propose another solution to non-linear state estimation called particle filters, which are also known as condensation or sequential Monte Carlo, which aim to dissolve the ambiguity created by occlusions and overlap by applying probabilistic models of the targets shape and motion. The strength of this method lies in its systematic treatment of non-linearity and non-Gaussianity. The algorithm proposed by Isard et al., is a fusion of a statistical factored sampling algorithm for static, non-Gaussian problems with a stochastic model for object motion. The result is an algorithm for tracking rigid and non-rigid motion that uses learned dynamical models, together with visual observations, to propagate the random set over time.

Breitenstein et al. [117] propose to extend the condensation algorithm with continuous confidence updates from pedestrian detectors and online trained, instance-specific classifiers as a graded observation model. Their main contribution is the exploration of how these unreliable information sources can be used for multi-person tracking in a particle filter framework. The resulting algorithm robustly tracks a large number of dynamically moving persons in complex scenes that contains occlusions. Their approach does not rely on background modelling, and operates entirely in 2D, thus requires no camera or ground plane calibration.

Lee et al. [118] compare the performance of non-linear Kalman filters and particle filters for a multi-vehicle flocking system using range measurements. The

authors use a distributed version of the Extended Kalman filter and a distributed Markov Chain particle filter, where the distributed implementation in both cases is done using consensus-type algorithms. The performance of the estimators are compared as the system complexity (number of vehicles) and measurement frequency are varied. The authors show that for simple systems (few vehicles) or high measurement frequency the Kalman filter method has lower expected error than particle filters, while for complex systems (many vehicles) or low measurement frequency the particle filter method is both more robust and more accurate.

Won et al. [119] presents an approach that combines Kalman and particle filters. The authors estimates position and orientation using one position sensor and one inertial measurement unit. The orientation is estimated using a particle filter and the position and velocity using a Kalman filter. The authors show that the orientation errors using the proposed method are significantly reduced compared to the errors obtained from using only Kalman filter. In addition the authors note that the proposed method can estimate orientation even in the presence of Gaussian position sensor noise.

To summarise, Kalman filters are an optimum observer that estimates the states of linear Gaussian state-space models. The Kalman filter and its variants and extensions are the most commonly used filtering techniques, however, when the model is highly non-linear or the noise is non-Gaussian, particle filters are more suitable. To reduce the computational complexity, Kalman and particle filters can be combined, where the linear Gaussian part of a system are solved using a Kalman filter and the remaining parts are solved using particle filters. Finally, Kalman filters have been shown to outperform particle filters when the system complexity is low and the measurement frequency high.

3.2.3 *Tracking algorithms*

Long-term tracking in real-world conditions is a challenging proposition - made more difficult by a multitude of external factors. The two main challenges for trackers are handling the similarity of appearance between the target and other objects in the scene, and the appearance variations of the target itself. Most vision-based trackers follow the *predict-match-update* paradigm, in which an internal representation of the target is relied upon. This representation is com-

pared to measurements extracted from incoming frames at predicted positions to estimate the most likely target location.

The representation encodes the appearance and the shape of a principal into the model state. In general, the representation is a trade-off between accuracy of the description and invariance. The state must be descriptive enough to discriminate the target principal in cluttered scenes and when faced with other principals, while allowing a certain degree of flexibility to cope with deviations in the appearance of the principal, caused by changes in scale, pose, illumination or partial occlusions.

Trackers need a method to propagate the state of the target over time. This requires the tracker to recursively use state information to form a trajectory for each tracked principal. This links different instances of the same target over time, requiring the tracker to have a strategy to manage these trajectories when targets appear and disappear from the scene. Thus a tracker must initialise a track for an incoming principal and terminate the trajectory associated with a disappeared target. An initialisation usually occurs at the image boundaries or at specific entry areas such as doors or entryways. Similarly, a trajectory must be terminated when the target leaves the field of view of the camera, or when the tracking performance degrades under a predefined level.

Trackers may use different representations of the state. At the most basic level, the state of the principal can be represented by points, where the tracker estimate the translation of the principal. This estimation can be performed using frame-to-frame tracking, for instance using the Lucas-Kanade tracker [101] and extensions [120], which forms the basis of many point trackers. Recent work is directed towards optimising performance of these methods. The Kanade-Lucas-Tomasi (KLT) tracker [121] locates good features by examining the minimum eigenvalue in a gradient matrix, and track these features using a Newton-Raphson method of minimising the difference between the frames. Takacs et al. [122] present a method that unifies tracking and video content recognition by introducing radial gradient transforms as a fast, high-quality feature descriptor. The authors combine tracking and recognition by using the same descriptors for both tasks.

Fukunaga et al. [123] propose the mean shift method as non-parametric feature-space analysis technique. It locates the maxima of a density function given a discrete data sampled from that function. Starting with an initial estimate x, let a

kernel function $K(x_i - x)$ iteratively determine the weight of nearby points for re-estimation of the mean $m(x)$, as follows:

$$m(x) = \frac{\Sigma_{x_i \in N(x)} K(x - x_i) x_i}{\Sigma_{x_i \in N(x)} K(x - x_i)} \tag{6}$$

where $N(x)$ is the neighbourhood of x. In image segmentation, Comaniciu et al. [124] use mean shift as the computational module to provide a robust feature space analysis. In later work by Elgammal et al. [125] and Duraiswami et al. [126] the authors explore the use of the kernel density estimation with the fast Gauss transform to increase the frame rate of the mean shift implementation.

Kalal et al. [64] [127] propose a tracker that reformulates the *predict-match-update* framework into the tasks: tracking, learning and detection. The tracker follows the object from frame to frame. The detector locates appearances that have been observed during tracking and corrects the tracker if necessary. The learning component estimates errors performed by the detector and updates it to avoid these errors in the future. For the tracking component the authors use a median-flow tracker, where the principals are represented by a bounding box, within which a sparse motion field is estimated.

3.3 SUMMARY

The smart devices used in positioning systems are prevalently based on wireless or vision-based devices. Wireless devices rely on hardware level communication between signal transmitters and measuring units where the location of users are determined based on signal measurements such as time of flight, angle, and signal strength. Vision-based devices are a diverse field and rely on detection, segmentation and feature extraction techniques to determine the position of users. Segmentation is the process of identifying and delineating objects in images or videos that share certain characteristics and feature extraction is the process of extracting the relevant parts.

The objective of tracking is to associate location information with the target principal over consecutive device readings. This association can be especially difficult when multiple principals are in the environment, when principals are

occluded, or when the quality of the readings are poor due to environmental factors.

Tracking methods must solve two basic problems: the motion problem, i.e., predicting the location of an element being tracked in the next frame, and the matching problem, i.e., identifying the element in the next frame within the designated search region.

Tracking combined with point-of-entry or continuous authentication offers a novel approach to calm authentication. Principals are tracked from the point of initial authentication to the point where authorisation is requested. This approach relies heavily on accurate tracking information, which may require considerable effort to obtain. Nevertheless, for location-based services, where the location of principals is already of interest, this approach offers a good balance between accuracy, security and usability.

Part III

MODEL

4

PERSISTENT AUTHENTICATION

Traditional authentication systems are considered persistent as they only provide point-of-entry authentication; after the initial authentication the validity of the user's authentication session is not considered. Conversely, sensor-based authentication methodologies are considered transient as they continuously monitor the user's context and can react to any changes that might affect the user's authentication session. In persistent authentication we combine the persistence of traditional authentication systems with the transience of sensor-based methodologies.

The goal in persistent authentication is to extend authentication from a single discrete event to a persistent session by utilising the contextual awareness provided by smart environments. Persistent authentication tracks users from the authentication at a point-of-entry system, to the point where the authentication is needed. The sensors in the environment are used to provide location information and to subsequently validate the identity of the principals by performing continuous authentication.

In this way, persistent authentication provides contextual awareness to location-based services and serve as an integrated protection measure, distributed directly in the facility. Persistent authentication ensures the secure provision of services and provides a mean to detect unauthorised entry and access. Persistent authentication is non-intrusive and the service provision is completely transparent to the users, who simply observe that services are available as they are needed.

At its core, persistent authentication combines data from an authentication system, a smart environment and a set of biometric experts. Based on this information the system tracks authenticated principals and provides authorisation information to a location-based service. An overview of the components in the persistent authentication system is given in Figure 4. The figure shows how the sensor data and initial authentication are used as input in the persistent authentication component, which then interacts with the biometric experts to provide a confidence measure of the relationship between a blob and the identity of the corresponding principal. Here the term *blob* is used to describe the logical rep-

resentation of a principal provided by the smart environment. For a vision-based system this could mean the image region containing the principal, for a wireless-based system this could mean the contour describing the transmitter, and for a pressure sensitive floor this could mean the area currently occupied by the principal.

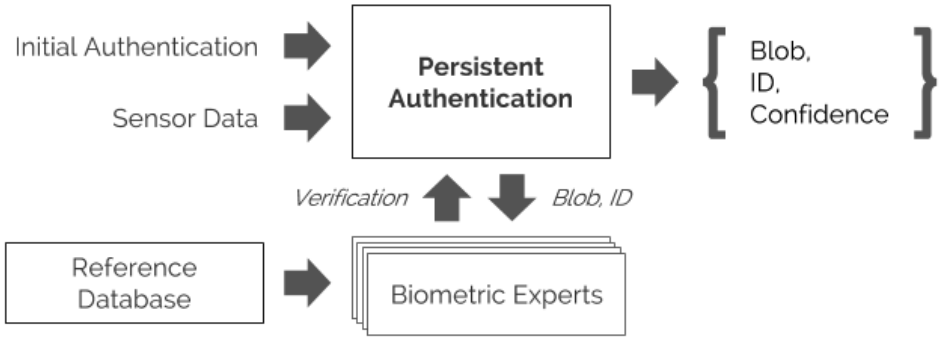


Figure 4.: Persistent authentication model

The process of providing the initial authentication of a principal is independent from the persistent authentication system. As such, the authentication can be performed by any authentication mechanism, including intrusive point-of-entry systems, biometrics or token-based systems. Once a principal has been authenticated the session is passed on to the persistent authentication system.

The smart environment provides the sensor data needed for tracking and continuous authentication of the principals. Conceptually, the persistent authentication model is independent from the sensor types used in the underlying smart environment, and requires only the location of principals and a method to periodically verify their identity.

The output of the persistent authentication component is the location of each blob in the scene, the identity of the corresponding principals and a confidence score, which describes the level of assurance that this assertion is correct.

4.1 AUTHENTICATION AND AUTHORISATION ZONES

In persistent authentication we define *authentication zones* and *authorisation zones* as the points where information regarding the principals transition from either the physical to the logical domain or vice versa. In the physical domain the zones represent the space in which principals interact with the authentication mechanisms and the location-based services. In the logical domain this is represented as the identity and credentials associated with each principal. The zones are defined such that the transition between the physical and logical domain can be performed reliably. In most cases, this is accomplished by restricting the physical space of these zones to hold only a single principal at the time, e.g., turnstiles used in the public transportation system and in large office buildings. Alternatively, additional sensors can be added to the smart environment, allowing the persistent authentication system to identify single principals, even in densely populated areas.

Figure 5 shows an example of how these zones can be defined and illustrates the authentication and authorisation process. A principal (grey/green dot) enters an authentication zone (dashed circle) and an authorisation zone (dashed square). The authentication and authorisation only succeeds if no other principals (red dots) are present in the zone.

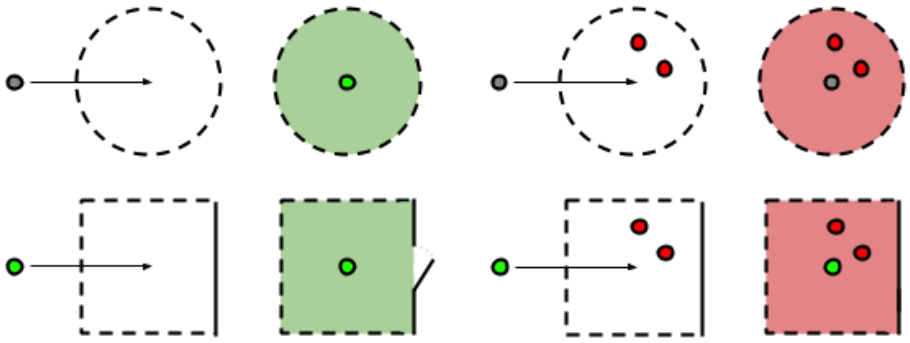


Figure 5.: Authentication and authorisation zones

4.2 MODEL AND WORLD STATE

In persistent authentication the physical representation of information is known as the world state, and concerns the location of principals in the environment, their identity, and any other security relevant situational information that is captured by the smart environment or used in the authentication process. The logical side is known as the model, and concerns the model representation of the physical domain and any appurtenant contextual information.

The distinction between the model and the world state can be seen as a statistical relationship between the measurements and the environment. In a camera, for instance, the three-dimensional world is projected into a two-dimensional set of measurements to form an image, which is a representation of the world. The objective of persistent authentication is to support mechanisms that use this representation of the world to make informed decisions that merges logical and physical information. However, there are two problems. First, the measurement process is noisy; what we observe is not a perfect representation of the world state but a noisy estimate. Second, the relationship between the world and measurements is generally many to one: many real-world configurations may produce the same measurements, e.g., two different principals may have the same appearance, speed or trajectory. We use probability to describe this relationship and to take noise in the data into account.

Let a continuous random variable y denote a real-world measurement that may be uncertain. If the random variable is observed over several instances $\{y_i\}_{i=1}^{I}$, it might take a different value on each occasion. Some values may occur more often than others, which is described by the probability distribution $Pr(y)$ of the random variable. The conditional probability of another random variable x given that y takes value y^* tells us the relative propensity of the variable x. The conditional probability is written as $Pr(x|y = y^*)$ and describes the relative probability that x takes various values after having observed $y = y^*$. For brevity, we will in the following write the conditional probability relation as $Pr(x|y)$, without explicitly defining the value $y = y^*$ to give a more compact notation. The notation and definitions introduced in the above and used throughout this thesis is presented as recommended by Price in [128].

The calculation of the conditional probability, $Pr(x|y)$, is done using Bayes' rule, which states the following:

$$Pr(x|y) = \frac{Pr(y|x)Pr(x)}{Pr(y)} \tag{7}$$

The left-hand side of Equation 7 is the a-posterior. It represents the probability of x given the observed measurement y. On the right-hand side, the prior, $Pr(x)$, is the initial degree of belief in x and the quotient $Pr(y|x)/Pr(y)$ represents the support y provides for x, where the numerator is known as the likelihood and the denominator known as the evidence.

In this thesis we use random variables, starting at the beginning of the alphabet, $a, b, c, ...$ to refer to principals in the model. A principal who authenticates using the i-th authentication zone, $Auth_i$, generates an authentication score s_i, which is the logical representation of the physical authentication. The confidence in the identity of a principal is given by the probability that the logical representation is equal to the physical representation and expressed as $Pr(a)$.

The confidence in the identity of a principal after receiving the authentication score s_i is expressed as the conditional probability $Pr(a|s_i)$, likewise the confidence in identity after receiving a biometric similarity score, y_i, is expressed as $Pr(a|y_i)$. This biometric score may be from the i-th biometric expert, $Expt_i$, or alternatively it may be a fused score that combines multiple biometric experts. Finally, we use e_i to express the noise that will invariably occur in some of the measurements, and we express the confidence in the identity of a principal after observing a noisy measurement in a similar fashion $Pr(a|e_i)$. Table 2 shows a summary of the notation used in the persistent authentication model.

Given an authentication score s_i for a principal a we can compute the confidence score $Pr(a|s_i)$ as the a-posteriori probability using Bayesian inference. The confidence score depends on the error rate associated with the underlying authentication factor, described by the likelihood $Pr(s_i|a)$, the confidence in the authentication model described by the model evidence $Pr(s_i)$, and the prior confidence $Pr(a)$. In the active stage of authentication, where the initial authentication of the user occurs, the prior is equal to the operational parameters of the chosen authentication system, whereas, in the passive stage, where continuous authentication occur, the prior is equal to the previous confidence score. Thus, given an authentication score of a principal we express Equation 7 using our definitions:

Table 2.: Notation used in the persistent authentication model

Description	Notation
Principals	$a, b, c, \ldots$
Authentication zone	$Auth_i$
Authentication score	s_i
Biometric expert	$Expt_i$
Biometric score	y_i
Event	e_i
Confidence in identity	$Pr(a)$
Confidence given authentication score s_i	$Pr(a\|s_i)$
Confidence given biometric score y_i	$Pr(a\|y_i)$
Confidence given noisy measurement e_i	$Pr(a\|e_i)$

$$Pr(a|s_i) = \frac{Pr(s_i|a)Pr(a)}{Pr(s_i)} \tag{8}$$

The confidence in the identity of a principal is adjusted based on environmental factors such as, noise, changes in illumination or occlusions. The variable e_i represents an uncertain or noisy measurement caused by an environmental factor and results in a decrease in the confidence score $Pr(a|e_i)$. Conversely, a positive biometric comparison y_i, will result in an increased confidence score given by $Pr(a|y_i)$. Both of these scores are expressed by substituting the corresponding terms in Equation 8.

To ensure the secure provision of location-based services, the persistent authentication model will revert to a failsafe state when tracking is lost or there is uncertainty in the identity of a principal. In these cases, $Pr(a)$ is either set to zero, thus effectively revoking the authentication session, or set to a fraction of the likely candidates, e.g., $Pr(a)$ would be set to 0.5 if two principals are equally likely.

Ultimately, the updated score reflects the system's confidence in the identity of the principal and determines the provision of location-based services.

4.3 ALGORITHM

In the following we present an overview of the persistent authentication algorithm. The persistent authentication system is initiated with the initial authentication of a principal as input. From this point on, the principal is tracked throughout the environment, using the sensors of the smart environment. Tracking is done by processing the sensor data and extracting the blobs corresponding to the principals. Probability is used to describe the confidence in the correspondence between the model and the world state, and thus the confidence in the identity of a principal.

The pseudo code for persistent authentication is presented in algorithm 1.

Input: Initial authentication $Pr(a|s_i)$
while *sensor data* **do**
  label blobs in the sensor data **throws** e_i
  track the position of principal a **throws** e_i
  if $Expt_i$ *generates similarity score* y_i **then**
    update confidence $Pr(a|y_i)$
  end
  if $Auth_i$ *principal in generates authentication score* s_i **then**
    update confidence $Pr(a|s_i)$
  end
  if *service requested* **then**
    return $Pr(a)$
  end
  catch e_i: **update** confidence $Pr(a|e_i)$
end

Algorithm 1: Persistent Authentication

Persistent authentication requires accurate labelling of the sensor data and a tracking methodology that can take noise and measurement errors into account. The challenge in persistent authentication is to generate a robust tracking score for all authenticated principals. This is made more difficult as the tracking scores must be generated in real-time, which requires very fast computations, often on embedded or limited hardware.

The first step of the loop, **label**, processes the sensor data and extracts the blobs corresponding to the principals in the scene. To ensure the accuracy and efficiency of the labelling, state of the art processing and segmentation techniques are considered. The labels are used in the next step of the algorithm, **track**, which solves the motion and correspondence problems for each detected blob in the scene. If an uncertain or noisy measurement occurs in either the **label** or **track** step, the algorithm throws an event, e_i, which is caught and the confidence score $Pr(a|e_i)$ updated accordingly.

In addition, the confidence score is adjusted when a principal enters an authentication zone and re-authenticates or when a biometric expert generates a similarity score for the principal. A re-authentication will generally result in a very high confidence score, as most point-of-entry authentication systems have a low or even negligible false acceptance rate. In contrast, a biometric authentication depends on the quality of the biometric sample and, as we do not impose any restriction on the capture of these samples, the result may only be a slight increase of the confidence score.

Finally, if service is requested the confidence in the identity of the principal is passed on to the location-based service, which can make an informed decision based on the situational awareness provided by the persistent authentication system.

The adjustments of the confidence score allows the system to operate in environments where occlusions, illumination changes and other noisy measurements affect the tracking accuracy, such that the confidence in the identity of a principal changes based on the quality of the tracking. An example is shown in Figure 6. The figure shows two paths, a solid line that corresponds to the motion of a principal a and a dashed lined that corresponds to the motion of principal b. Events on the paths have timestamps, and the time t_0 corresponds to the initial authentication, where a authenticate using a point-of-entry authentication mechanism, giving an initial confidence of 1.

Both principals are reliably tracked from the point of initial authentication until the time t_1, where the occlusions of the principals causes uncertainty in which of the paths the tracked principal a is following. As a result, the confidence in the identity of a is lowered. The magnitude of this decrease in confidence depends on the evaluation of $Pr(a|e_i)$, but for the sake of the example, we assume that there is an equal chance of a following either path.

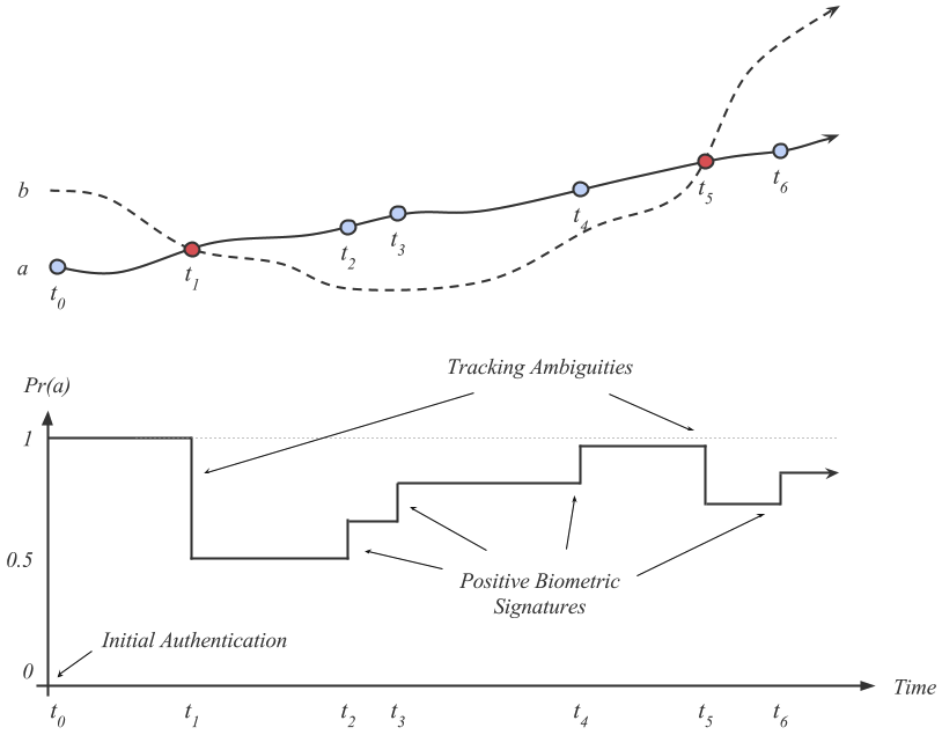


Figure 6.: The confidence in a's identity decreases when the paths of a and b intersect and increases with positive biometric signatures.

The remote biometrics of the principals are continuously measured and at time t_2, t_3 and t_4 positive signatures for a are captured. The resulting biometric evaluation is used to increase the confidence in the identity of a. The increase in confidence depends on the quality of the biometric sample and the output of the biometric experts. This cycle of decreasing confidence due to noise or occlusions and increasing confidence with positive signatures continues as long as a is in the environment.

4.4 SECURITY POLICIES

In all critical environments there is a need to limit access to resources and to place restrictions on the availability of services. These limitations and restrictions are formalised into security policies that indicate who are allowed access or service. One of the most interesting applications for location-based services are the development of location-based access control. Access control represents a key aspect that can be radically changed by the availability of location information. Verifying the location of a principal in addition to the checks performed by traditional authentication methods can greatly improve the security of a facility.

The location-based conditions that are particularly interesting for access control policies are the following [129]:

- *Position-based conditions* related to the location of the users, e.g., the evaluation of whether the user is in a certain area or in a specific room in a building, or alternatively, if the user is in the proximity of other users.
- *Movement-based conditions* related to the mobility of the users, e.g., their velocity, acceleration, or direction.
- *Interaction-based conditions* related to the relationships between multiple users or entities, e.g., the number of users within a given area.

The first step to support location-based access control is to identify how location information is queried and what kind of response the location-based service

returns. Traditional location-based services [130] usually model this as functions of the form:

$$predicate(parameters, value) \rightarrow [return\ value, confidence, timeout] \quad (9)$$

This function states that the evaluation of a location predicate over the parameters has the stated value. The result is a return value with a corresponding confidence and timeout. The confidence value expresses the level of reliability associated with the assessment, which is given by the confidence in the identity of the principal in the persistent authentication model. Finally, the timeout represents the time validity of the result.

For the sake of simplicity, we consider the queries in our model to return a boolean value. For instance, a query can evaluate whether a user is located in an authorisation zone, where the result (true or false) and the confidence in this assessment are used for service provision.

Table 3 presents a few examples of specific predicates that utilise location-based information. The predicates *inarea* and *disjoint* evaluate the location of the users with regards to a given area. The areas are defined as a geometric model that describe either the vicinity of the user or refer to specific entities such as authentication zones, rooms in the building, etc. The predicate *distance* and *velocity* evaluates the position and mobility of the users and *density* and *local density* evaluate the spatial relationships between users in a given area. For a full overview of the research issues and emerging trends related to location-based information we refer to work by Bettini et al [129].

4.4.1 *Location-based Access Control*

The location-based predicates are used in conjunction with persistent authentication to allow physical security policies to be described logically. An example of a common security policy concerns the physical security of a restricted area, protected with an access control mechanism.

We formulate such a policy using persistent authentication and the location-based predicates by expressing the provision of a location-based service at the j-th authorisation zone, $Atho_j$, as the function S_j. The evaluation of the function

Table 3.: Examples of location-based predicates

Predicate	Description
inarea(user, area)	Evaluate whether a user is located within the given area.
disjoint(user, area)	Evaluate whether a user is outside a given area.
distance(user, entity, min, max)	Evaluate whether the distance between the user and entity is within the given interval.
velocity(user, min, max)	Evaluate whether the user's speed falls within the given range.
density(area, min, max)	Evaluate whether the number of users in the area falls within the given interval.
local_density(user, area, min, max)	Evaluate the density within a relative area surrounding user.

$S_j(a)$ uses the location-based predicate $inarea(a, j)$ to define the area of the authorisation zone. The result of the evaluation is a binary decision, accept or reject, depending on whether the confidence returned by the predicate *inarea* exceeds a service provision threshold Δ_j. These service provision thresholds Δ_j are defined for the location-based services and authorisation zones to dictate the level of confidence required for service provision. The notation is summarised in Table 4.

Table 4.: Notation used for service provision in persistent authentication

Description	Notation
Authorisation zone	$Atho_j$
Service provision function	S_j
Service provision threshold	Δ_j

The evaluation of the function $S_j(a)$, for a location-based service at the j-th authorisation zone, is thus defined as follows:

$$S_j(a) = \begin{cases} accept & Pr(a) > \Delta_j \\ reject & otherwise \end{cases} \tag{10}$$

As an example, consider a physical security policy concerning a location-based service represented by a context-aware access control mechanism. To define this policy using the above notation we present the following criteria:

> *Authentication:* Principals that enter the authentication zone $Auth_i$ and present their credentials, generate authentication scores $\{s_i\}_{i=1}^{I}$. The principals are authenticated and their privileges associated with the sessions. If no credentials are presented the principals remain unauthenticated.
>
> *Privileges:* An access control list is used to determine the privileges of the principals. If the principals exit the surveilled area, or if they are lost/occluded and the tracking algorithm is unable to re-associate the session, then the authentication session is rendered void.
>
> *Access:* Principals are granted access to the restricted area if the correct privileges are associated with their session and they are fully within the authorisation zone $Atho_j$ and the confidence $Pr(a)$ meets the threshold defined in the service provision function $S_j(a)$. In addition, an evaluation of *density(j, 1, 1)* is used to ensure that if any unauthorised principals are present in the zone, access is denied.

While this security policy is quite simple, it suffices to illustrate the advantage of persistent authentication in smart environments. Persistent authentication solves a common problem that exists with physical security policies, namely, that in practice these policies are very difficult to maintain. For example, with traditional access control mechanisms, authenticated principals can allow unauthorised people into the restricted area - either willingly or as a result of intruders that tailgate the authenticated principals. Persistent authentication solves this problems by evaluating the *density* predicate in the authorisation zones and if any unauthorised principals are detected in the zone, access is denied. This ensures that only authenticated principals are allowed access into the restricted area and prevents tailgating. Conversely, if the usability of the system are of concern, the security policy can be modified to allow multiple authenticated principals to enter the restricted area at the same time. This speeds up the authorisation time and increases the usability of the access control system.

Rejection of service may occur, either because the confidence in the identity of the principal is not high enough to meet a given service provision threshold, Δ_j, or because the authentication session have been completely revoked. In these cases the principal has to re-authenticate at the nearest authentication zone to receive further services. It must be emphasised that this will be a rare occurrence if the tracking methodology is robust. Moreover, it is possible to place authentication zones at several strategically selected locations to lessen the inconvenience of re-authentications.

The flexible design of persistent authentication makes it easy to change or replace parts of the policy. For instance replacing the authentication mechanism (*Authentication*) and the access control mechanism (*Privileges* and *Access*) to reflect other location-based services. Similarly, the *Access* policy can be changed, for instance, to allow authenticated principals to escort guests into the protected area.

In this way, the system acts like a sensor enhanced access control system [131], where persistent authentication has the role of the context manager. This setup results in a fine-grained and flexible access control mechanism that provides situational awareness to a location-based service and takes both the logical entities and the physical relationship between principals and environment into consideration.

4.4.2 *Virtual Walls*

As an interesting aside, persistent authentication allows the enforcement of physical access control policies based on the concept of *virtual walls* [132]. Virtual walls extends the protection provided by physical walls into the virtual realm. For instance, we intuitively assume that if we lock the front door of our home it is safe from intruders. Thus, we rely on the physical security offered by the confines of our home, and assume the walls of our homes to be impenetrable to all, but the most enduring intruders. With persistent authentication we can extend this intuitive notion of security by introducing virtual walls in the environment.

Virtual walls are particularly interesting in cases where traditional access control is difficult or even impossible to implement. With the rise of open plan offices we witnessed a shift in paradigms to focus on open, accessible, spaces where

workers, visitors and inhabitants are free to roam and interact with each other. Whilst it can be argued that open place offices foster more communication between staff and boost the community spirit then, from an access control point of view, a new approach is needed to ensure the same level of protection as offered by traditional layouts.

Consider the floor plan shown in Figure 7. This illustrates a typical open plan office, with a reception, a waiting area, some private offices and a conference room. After entering the premise there is no additional access control. Hence, it is up to the staff and the receptionist to ensure visitors do not gain unwanted entry into the private areas of the office.

In persistent authentication we propose a policy framework that introduces virtual walls in an environment to provide location-based access control. We define virtual walls as follows: A virtual wall $w_i = \langle area, \{users\}, response \rangle$ covers a given *area* and optionally concerns a list of *users*. Based on this information a *response* is formed, indicating the action required of the underlying system.

The area is defined as a geometric model that describes specific entities, such as the waiting area or the conference room as seen on Figure 7. The list of users can be defined for each virtual wall or obtained from an access control list. Finally, the response is used to indicate the action required of the underlying system. This response is mapped to actions such as *unauthorised access*, or *occupancy limited reached* (for instance due to fire regulations). This adds flexibility, and allows a finer control over the dissemination of the virtual walls.

Virtual walls, in contrast to physical walls, do not provide or enforce any physical security. Thus, in essence virtual walls are a means to raise specific actions required of the underlying systems to ensure the physical security of a premise. For instance raising alarms if unauthorised people are to enter one of the private offices in Figure 7, or notifying the receptionist if visitors do not wait for an escort in the waiting area.

4.5 DETECTING INTRUSIONS AND HOSTILE RECONNAISSANCE

Detecting illegitimate access and hostile reconnaissance is of interest to any security system. In the persistent authentication framework we conjecture that such attacks are distinguishable from normal patterns in a suitable feature space, but

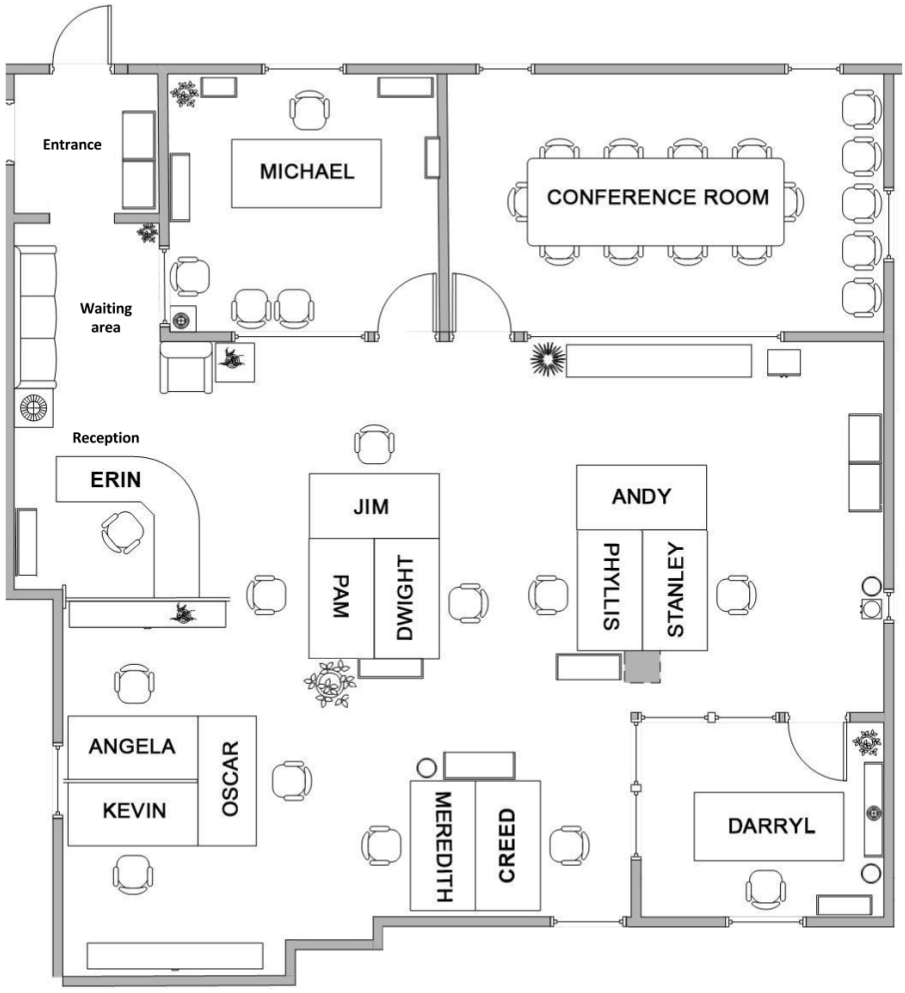


Figure 7.: Example of an open plan office

detection is difficult as the malicious adversaries often seek to mask their behaviour to appear normal.

While adversaries mask their behaviour before attacks, then the occurrence of intrusions, infiltration, bombings, robberies, etc., create noticeable disruption, either as an objective or as a byproduct. Hence, it is trivial to identify attack patterns in the dataset post incident. It is then possible to use the identified patterns as a labeled dataset to either define heuristics or as training data for machine learning techniques. The drawback of this approach is that only known incidents can be leveraged in detection and that the labeled dataset needs to be manually revisited with each new attack, which is an ineffective and expensive procedure.

Furthermore, false positives and negatives may occur if the training dataset is insufficient for the chosen application, or if the dataset is lacking some aspect of normal activity. These errors are further compounded if the chosen evaluation methods or heuristics are not well defined. Thus detecting intrusions and hostile reconnaissance in the persistent authentication framework requires additional information regarding the behaviour of principals and their interactions with the environment.

As a result, we integrate our efforts in persistent authentication with the analysis of architectural morphology. Architectural morphology is part of the general field of infrastructure analysis and can be understood as the analysis of interconnected spaces. The general idea in this field is based on the rationale that a space can be analysed as a series of components that form a network of choices. These choices can then be represented as maps and graphs that describe the relative connectivity and integration of the space.

This approach of understanding space as a set of relations described, limited, or promoted by architecture is deeply rooted in the basic notion that spatial structures contain social structures and vice versa. These structures promote patterns in people's behaviour, with regards to their relation to environment. When people move through familiar environments their motion is goal oriented. Typically, people in public buildings walk between a finite number of points of interest (doors, corridors, etc.) often following set paths that are determined by a combination of practicality and unwritten social rules. In addition people generally adhere to the general closeness measure, which means they take the shortest route to their destination. Conversely, in private and more organised spaces, occupation and

movement are dominated by employees, which instead conform to the organisational and functional layouts that are mediated through the space.

From a security perspective the arrangement of space in an organisation should ensure that there are limited opportunities to overview sensitive areas of the operations, and from where to undisturbed and unrecognised perform or plan attacks. Without a structured public-to-private interface, detecting intruders and illegitimate access becomes considerably harder, and adversaries performing hostile reconnaissance can more easily map behaviour and habits of personnel, managers, and other sensitive resources of the organisation. Thus the arrangement of space in an organisation should ensure the spatial resilience of the building and reduce the number of points from which reconnaissance can be performed.

The points that may be targeted for hostile reconnaissance includes (1) crowded locations, (2) locations through where movement is guided, (3) locations of important functions or service interactions, and (4) locations of internal and/or private operations potentially exposed to the public or publicly accessible spaces.

The arrangement of public space should integrate these points and structure the network of choices such that visitor flows form recognisable patterns. This makes points of potential reconnaissance more easily detected and deviations from the patterns more easily recognised.

The process of identifying deviations are known as anomaly detection and refers to the problem of finding patterns in data that do not conform to expected behaviour. While anomaly detection is widely used in multiple fields, most people with a computer science background will associate anomaly detection with network intrusion detection systems (IDS). An IDS monitors network traffic and classifies it as being either normal or anomalous. This classification can be based on heuristics, rules, patterns or predefined signatures. The anomalies in the data indicate the presence of significant, actionable, information, e.g., if an IDS detects an anomalous traffic pattern it may mean that a computer on the network has been compromised.

Anomaly detection is straightforward on an abstract level; define a region representing normal behaviour and classify any observation in the data which does not belong to this region as an anomaly. However, several factors make this apparently simple approach very challenging. Chandola et al. [133], outline the following challenges in anomaly detection:

- Defining a normal region which encompasses every possible normal behaviour is very difficult. In addition, the boundary between normal and anomalous behaviour is often not precise. Thus an anomalous observation which lies close to the boundary can actually be normal, and vice versa.
- When anomalies are the result of malicious actions, the malicious adversaries often adapt to make the anomalous observations appear like normal, thereby making the task of defining normal behaviour more difficult.
- In many domains normal behaviour keeps evolving and a current notion of normal behaviour might not be sufficiently representative in the future.
- Availability of labeled data for training and validation of models used by anomaly detection techniques is usually a major issue.
- Often the data contains noise which tends to be similar to the actual anomalies and hence is difficult to distinguish and remove.

Due to these challenges, the anomaly detection problem is not easily solved and instead it is necessary to formulate the problem to include a set of assumptions. These assumptions may include factors such as the nature of the data, the availability of labeled data, the type of anomalies to be detected, etc.

Depending on the assumptions, the re-formulated problem may be solved statistically or using data exploration techniques. Statistical approaches may use heuristics and semi-supervised learning, whereas data exploration techniques process and structure the acquired data in different ways to visualise the different normal classes and detect the prevalent patterns to reveal anomalies in the data.

In the persistent authentication framework we focus on data exploration techniques to detect anomalies. We use data exploration techniques to detect patterns in the movement and flow of people in the environment. These patterns are used to detect people that exhibit unusual behaviour or behaviour that does not conform to the spatial configuration of the facility. We combine the data exploration with the evaluation of virtual walls as discussed in subsection 4.4.2. We define a set of virtual walls, based on the parameters of the location-based system, e.g., using the location and authentication session of the principals, to detect undesirable situations occurring in the building.

To support the evaluation we develop a dedicated tool that is capable of processing and structuring very large quantities of data, while taking the spatial configurations into consideration. To visualise the prevalent patterns in the data we integrate our findings with state of the art data visualisation tools to present the information in a meaningful way.

4.6 PRIVACY AND ETHICS

The pervasive nature of smart environments combined with the technological advances such as identification and tracking have the potential to disrupt the balance between the need for context-aware systems and the privacy of individuals. Here privacy is used to describe how far one can intrude into the personal affairs of an individual. As Alan Westin defines it, privacy is *"the desire of people to choose freely under what circumstances and to what extent they will expose themselves, their attitude and their behaviour to others"* [134], [135].

Delivery of location-based services require tracking of principals in the environment, often with the use of camera-based sensors, and raises a number of privacy and ethical challenges. Location information has the potential to allow an adversary to physically locate a person. As such, the principals have legitimate concerns about their personal safety, if such information should fall into the wrong hands. In addition, the identification of patterns in the behaviour of the principals can compromise the privacy of the inhabitants if the profiles are revealed to unintended users. Therefore, it is important that all sensitive information is revealed only on the need-to-know basis.

In the persistent authentication framework we seek to protect against any unnecessary privacy intrusions. We recognise that it is impractical to obtain informed consent from all observed principals, and thus have to accept some risk to privacy. We seek to only accept minimal, proportional and necessary risks and we shall at all times minimise these risks using all methods at our disposal, including but not restricted to:

- The anonymisation of individuals through not collecting or storing personal details. It should be noted that in cases where informed consent can be obtained from the principals, then anonymisation of individuals are not necessary and the collection and storage of personal information allowed.

- Using the minimal sensor resolution necessary to conduct our research.
- When possible and practical, transforming the data captured from the sensor into alternative formats, such as spatial representations.
- Deleting the original sensor recordings as soon as the data has been transformed or, alternatively, when the recordings are no longer necessary to hold.
- Any collection of data will conform to any and all national legislation affecting such collection, and it is the responsibility of those collecting or using such data to ensure they have made themselves aware of such legislation and are adhering to them at all times.
- Any data collected that is transferred outside of the EU, including to partners of the RIBS project situated outside the EU, must comply with relevant EU rules and international/bilateral agreements incorporated into EU law. Data collection will comply with relevant national and EU legislation on data storage, collection and privacy as well as the Fair Use Principles.

As a point of departure we consider the application of the Danish laws and regulations to our research. We refer interested readers to the international survey on privacy and data protection by Banisar et al. [136], which gives an overview of the differences in national legislation and regulation.

In Denmark the right of privacy is recognised in the constitution and there exists a regulatory body to oversee and enforce that the regulations are implemented and upheld. In addition to the constitutional rights, the main legislation in the area are *The Act on Processing of Personal Data (Act No. 429 of 31 May 2000)* which entered into force on 1 July 2000. The act implements Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. By only accepting minimal, proportional risks using the methods outlined above, the Danish Data Protection Agency found persistent authentication to comply with all rules and regulations. As a result, persistent authentication is approved for installation in all private premises and requires only proper signage indicating the operation.

Further, the RIBS project encompasses a truly multidisciplinary approach to research incorporating engineers, natural scientists and social scientists, all work-

ing in disparate fields and from both academia and industry. As such there is no single research or professional code of practice which is bespoke to such encompassing research.

However, it is acknowledged that all those involved in the RIBS project should be conducting their research in accordance with the highest ethical codes of conduct. It is envisioned that all those involved with RIBS are operating under suitable ethical codes in this regards. Academics from the various universities involved will be operating under their institutions' ethical codes of research practice and thereby bound by them as well as bound by their professional codes of practice.

4.7 SUMMARY

Persistent authentication presents a new approach to authentication that enables the secure provision of location-based services through calm authentication of mobile users in a smart environment. The objective is to shift the current authentication paradigm from a single discrete event to a continuous session. This is accomplished by utilising the contextual awareness provided by the smart environment to track principals from the point of initial authentication to the point where authorisation is requested by the location-based service.

Authentication zones and authorisation zones are defined as the points where information regarding the principals transition from either the physical to the logical domain or vice versa. This transition is a noisy process, thus probability is introduced to describe the relationship between the persistent authentication model and the measurements.

Data exploration techniques are considered to detect patterns in the behaviour of the occupants and identify principals that exhibit unusual behaviour. Virtual walls, based on the parameters of the persistent authentication system, are considered to raise alarms if undesirable situations are occurring in the building.

In the persistent authentication framework we seek to protect against any unnecessary privacy intrusions by only accepting minimal, proportional risks. As a point of departure we consider the Danish laws and regulations and their applications to our research.

5

ERROR-RATE-BASED FUSION

In this chapter we present *Error-Rate-based Fusion*, a novel fusion strategy that transforms the output of individual biometric experts into objective evidences and combine them using Bayesian inference.

In more details, let us assume that a biometric expert generates a similarity score y_i and the expert makes the decision that the claimant is genuine. This decision is not completely reliable, as the false acceptance rate of the expert, given the similarity score y_i, represents the probability that the claimant is an impostor. Similarly, the false rejection rate for the given score represents the intrinsic probability of incorrectly rejecting a genuine user. In our fusion strategy we use Bayesian inference to combine these false acceptance and false rejection rates from different scores, calculated by different experts, to generate a confidence value representing the probability that the claimant is genuine.

For biometric verification we consider two class labels, A and $\bar{A}$, where A is assigned when the expert concludes that a claimant is genuine, and $\bar{A}$ is assigned if the authentication status of the claimant is unknown. Note, that we use capital letters to distinguish the class labels from the notation used to represent confidence scores in the persistent authentication framework.

If the claimant is A but the expert wrongly labels him $\bar{A}$ then this event is called a false rejection (FR). Similarly, if the claimant is not A and an expert wrongly labels him A then this event is called a false acceptance (FA). The false acceptance and the false rejection rates (FAR and FRR) are calculated as the fractions of, respectively, the false acceptances and the false rejections taken over all the events.

5.1 OPERATIONAL PHASES

We distinguish between three operational phases of a biometric expert. The first phase is called the development phase, in which an expert operates on a development dataset. The development dataset consists of a large number of biometric

samples and the corresponding true class labels. The purpose of the development phase is to optimise the values of different operational parameters within the expert algorithm, and thus the phase is not directly relevant to our fusion strategy.

The second phase is called the training phase, in which an expert is evaluated by an evaluator on a training set, which does not contain the true labels of claimants, however, the true labels are known to the evaluator. The purpose of the training phase is to determine a-priori errors in the experts decisions, namely a-priori FAR and FRR. This phase is essential for our fusion strategy, as our strategy depends on these a-priori FAR and FRR values.

The third phase is called the test phase, in which a biometric system containing multiple experts is tested on a test dataset. This phase models the working of the biometric experts in an actual operating environment. The FAR and FRR values that occur in this phase are called the a-posteriori FAR and a-posteriori FRR. It is important that the training dataset does not overlap with the development and the test datasets, because the derivation of FAR and FRR assumes the independence of the training dataset.

5.2 FORMAL MODEL AND NOTATIONS

Let us consider N biometrics experts. The output of the i-th expert is a similarity score, $y_i \in \mathbb{R}$, where $1 \leq i \leq N$. For a decision threshold Δ_i, the decision function is defined as follows:

$$decision(\Delta_i, y_i) = \begin{cases} accept & \text{if } y_i \geq \Delta_i \\ reject & \text{otherwise} \end{cases} \tag{11}$$

For the training set of the i-th expert, let $\{y_i^A\}_{i=1}^I$ be the number of A's samples, and let $\{y_i^{\bar{A}}\}_{i=1}^I$ be the number of the samples that are not from A. The size of the training set is $\{y_i^A\}_{i=1}^I + \{y_i^{\bar{A}}\}_{i=1}^I$. For a given value of Δ_i, the false acceptance rate and false rejection rate are defined as follows:

$$FAR(\Delta_i) = \frac{\text{Number of FA at } \Delta_i}{\{y_i^{\bar{A}}\}_{i=1}^I} \tag{12}$$

$$FRR(\Delta_i) = \frac{\text{Number of FR at } \Delta_i}{\{y_i^A\}_{i=1}^I} \tag{13}$$

In a standalone operation, where there is only one expert and no need for score fusion, Δ_i is set to a fixed value, such that the cost of errors is minimum, e.g., if both FAR and FRR are of equal concern then Δ_i is set to the equal error rate (EER).

In our fusion strategy, we are interested in a combined decision of N biometric experts after fusion, and for this purpose we convert the similarity score y_i into the equivalent FAR and FRR values.

With the similarity score y_i, let the functions $FAR(y_i)$ and $FRR(y_i)$ be the false acceptance rate and false rejection rates of the i-th expert with $\Delta_i = y_i$. Since $y_i \in \mathbb{R}$, these functions are continuous, such that $FAR(y_i) \in \mathbb{R}$ and $FRR(y_i) \in \mathbb{R}$. The parametric way of modelling these functions is to find an analytical expression, such as a Gaussian distribution. Usually, $FAR(y_i)$ and $FRR(y_i)$ are not as smooth as a Gaussian distributions, and an approximation with simple analytical expressions will introduce estimation errors.

For precise evaluation of $FAR(y_i)$ and $FRR(y_i)$, we use a non-parametric approach, and model them as step functions, in which Δ_i can only take m different values: $\Delta_i \in \{\delta_i^1, \ldots, \delta_i^m\}$, where $\delta_i^1 < \cdots < \delta_i^m$. We call these values of Δ_i error decision thresholds (EDTs). This means that $FAR(y_i)$ and $FRR(y_i)$ are defined over a set of m EDTs:

$$FAR(y_i) = \begin{cases} FAR(\delta_i^m) & \text{if } y_i \geq \delta_i^m \\ FAR(\delta_i^{m-1}) & \text{if } \delta_i^m > y_i \geq \delta_i^{m-1} \\ \ldots & \ldots \\ FAR(\delta_i^1) & \text{if } \delta_i^2 > y_i \geq \delta_i^1 \\ 1 & \text{Otherwise} \end{cases} \tag{14}$$

$$FRR(y_i) = \begin{cases} FRR(\delta_i^m) & \text{if } y_i \geq \delta_i^m \\ \ldots & \ldots \\ FRR(\delta_i^1) & \text{if } \delta_i^2 > y_i \geq \delta_i^1 \\ 0 & \text{Otherwise} \end{cases} \tag{15}$$

The different values of Δ_i are illustrated in Figure 8, with a typical plot of the probability density functions (PDF) of expert scores. The figure illustrates that the similarity scores for a genuine user are distributed on larger values as compared to that of an impostor. The figure also shows the point of equal error rate, where the false acceptances and false rejections have the same value.

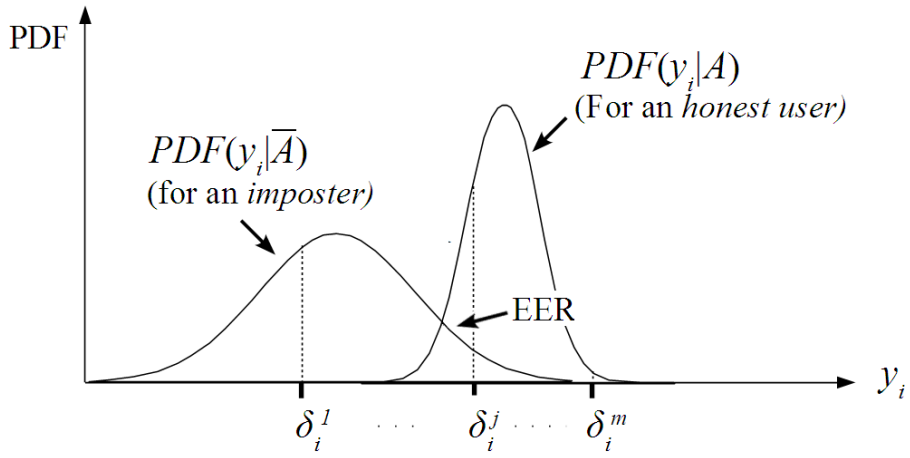


Figure 8.: Error Decision Thresholds (EDTs) and Probability Density Functions (PDF) of typical expert scores.

We determine $FAR(y_i)$ and $FRR(y_i)$ from the training set on each value of Δ_i. As y_i falls somewhere in the range of an error decision threshold, then the step functions over approximate actual FAR and FRR values, which, although a pessimistic approach, is sound from a security perspective.

To illustrate an error-rate-based fusion system, consider a verification system that contains N biometric experts. When biometric data of a claimant is available from the sensors, the system invokes the experts with the claimed identity A. Each expert extracts the relevant biometric features from the data and compares the extracted features with the biometric templates of A. Each expert then generates a similarity score y_i, and we compute $FAR(y_i)$ and $FRR(y_i)$ and fuse the similarity score with Bayesian inference. As our approach fuse the FAR and FRR values that have the same meanings across the different experts, there is no need for pre-fusion normalisation.

As shown in Figure 9, when the i-th expert generates a similarity score y_i, we compute $FAR(y_i)$ and $FRR(y_i)$ and fuse them based on Bayesian inference. The inputs from each of the N experts can be processed sequentially. This is due to the fact, that the Bayesian inference is invariant to whether the similarity scores

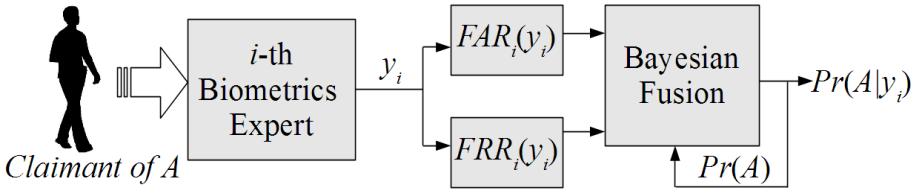


Figure 9.: Similarity scores are converted to equivalent FAR and FRR measures.

from all experts are processed in parallel or in a sequence. Similarly, the order of experts in the sequential processing does not affect the fusion process.

The system has an a-priori confidence that the claimant is A, which is represented as the probability measure, *Pr(A)*. The complementary confidence that the claimant is not A is *1 - Pr(A)*. We compute a-posteriori confidence, $Pr(A|y_i \geq \Delta_i)$, i.e., the probability that the claimant is A after receiving the evidence y_i that meets the decision threshold Δ_i. For brevity, we do not include the decision threshold in the probability expressions, and therefore we write $Pr(A|y_i \geq \Delta_i)$ as $Pr(A|y_i)$. The value of the a-posteriori confidence is computed as follows:

$$\begin{aligned} Pr(A|y_i) &= \frac{Pr(y_i|A)Pr(A)}{Pr(y_i)} \\ &= \frac{Pr(y_i|A)Pr(A)}{Pr(y_i|A)Pr(A) + Pr(y_i|\bar{A})Pr(\bar{A})} \\ &= \frac{Pr(y_i|A)Pr(A)}{Pr(y_i|A)Pr(A) + Pr(y_i|\bar{A})(1 - Pr(A))} \end{aligned} \tag{16}$$

On the right hand side of Equation 16, $Pr(y_i|A)$ is the probability that the i-th expert generates $y_i \geq \Delta_i$ if the claimant is indeed A. This is determined by the false rejection rate, $FRR(y_i)$, i.e., the probability that the i-th expert generates a score that is less than y_i even though the claimant is A. Therefore, we have:

$$Pr(y_i|A) = \int_{y_i}^{\infty} PDF(y_i|A) = 1 - FRR(y_i) = TAR(y_i) \tag{17}$$

Here, TAR is the true acceptance rate. Similarly, $Pr(y_i|\bar{A})$ is the probability that the i-th expert generates $y_i \geq \Delta_i$ if the claimant is not A, which is exactly the probability of a false acceptance. Therefore, we have:

$$Pr(y_i|\bar{A}) = \int_{y_i}^{\infty} PDF(y_i|\bar{A}) = FAR(y_i) \tag{18}$$

The relationship between the PDF of y_i, FAR, FRR, and TAR from Equation 17 and Equation 18, is illustrated in Figure 10.

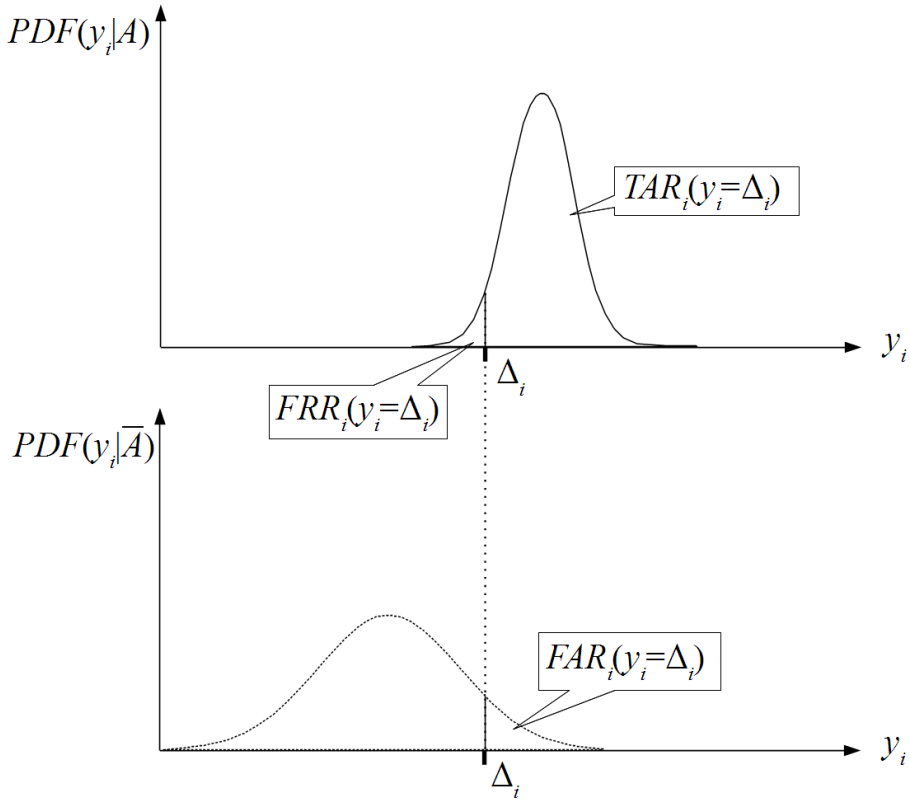


Figure 10.: The Areas of FAR, FRR, and TAR.

Next, we write Equation 16 in terms of FAR and TAR for $Pr(A|y_i)$ and $Pr(\bar{A}|y_i)$:

$$Pr(A|y_i) = \frac{TAR(y_i)Pr(A)}{TAR(y_i)Pr(A) + FAR(y_i)(1 - Pr(A))} \tag{19}$$

$$Pr(\bar{A}|y_i) = \frac{FAR(y_i)Pr(\bar{A})}{FAR(y_i)Pr(\bar{A}) + TAR(y_i)(1 - Pr(\bar{A}))} \tag{20}$$

Error-rate-based fusion uses Equation 19 and Equation 20 to fuse the outputs of multiple biometric experts, by taking into account the prior confidence level. As the scores are converted to probabilities these can be freely combined using simple operators or used as a-prior information in further processing.

To get an intuitive feeling of these equations, let us consider a traditional verification expert, which is assumed to be error free, e.g., a password-based authentication of claimant A, on a computer terminal. If the password is correct then the computer has full confidence that the claimant is A. For such an expert, the values of FAR and FRR are assumed to be zero. As expected, the confidence is evaluated to *1* in Equation 19 independent of a-priori confidence. In fact, any expert for which the value of FAR is zero will generate the confidence value of *1*, which is consistent with the fact that with zero false acceptances no impostor can ever be accepted by the expert.

5.3 EFFECT OF QUALITY ON FUSION

In practice, not every sample of biometric modalities are of the same quality. This may be caused by variation in environmental conditions or the relative position of a remote biometric sensor with respect to the user. The quality of a biometric signal, however, is non-discriminatory [137], which means that a difference in quality is not useful to distinguish between A and $\bar{A}$.

On the other hand, quality does affect the reliability of samples. In a parallel fusion approach, a common approach is to give more relative weight to high quality samples as compared to low quality samples. Since we process samples sequentially, this approach is not suitable to our error-rate-based fusion strategy. In the following, we discuss two approaches to include the effect of quality in our fusion strategy.

- The simplest approach to deal with poor quality samples is to discard such samples. In this case, an expert generates a quality score q_i, besides a similarity score, and if the quality score is below a minimum threshold, determined experimentally from the training set, then the biometrics sample is rejected.

- Because FAR and FRR vary with quality, we can use quality dependent, composite EDTs: $\{(\delta_i^1, q_i^1), \ldots, (\delta_i^m, q_i^1)\}, \ldots, \{(\delta_i^1, q_i^n), \ldots, (\delta_i^m, q_i^n)\}$, where $q_i^1, \ldots, q_i^n$ are n different quality thresholds. As a result, the functions $FAR(y_i, q_i)$ and $FRR(y_i, q_i)$ can be determined from the training set for each pair (δ_i^j, q_i^k), for $1 \leq j \leq m$ and $1 \leq k \leq n$.

However, under a set of reasonable assumptions, error-rate-based fusion is robust to change in quality, without explicitly considering the quality measure. In the rest of this section, we analyse why this is the case.

For error-rate-based fusion, we expect that the quality of a biometric sample should determine the change in the confidence on the identity of A, i.e., the increase in $Pr(A|y_i)$ due to a good quality sample of A should be more than the increase due to a relatively poor quality sample of A. Further, if the quality is too poor then no update in $Pr(A|y_i)$ should be made.

In this regard, an important observation is that degradation in quality adversely affects similarity scores [47]. Thus the similarity score between a good quality biometric sample and the template is higher than the similarity score between a poor quality biometric sample and the template. Therefore, for the EDTs $\{\delta_i^1, \ldots, \delta_i^m\}$, where $\delta_i^1 < \cdots < \delta_i^m$, a poor quality sample is likely to fall under the range of a lower EDT.

We also note that the quality of biometric samples greatly affect TAR, but it has a small effect on FAR, which is also evident in literature [46] [39]. As a result, it is unlikely that a bad quality sample from an adversary matches the template of A. Similarly, it is also unlikely that a bad quality sample from A improves the similarity score between the sample and the template of A.

Let the similarity score resulting from a good quality sample of A be y_i, and let the similarity score resulting from a relatively poor quality sample of A be y_i'. Based on our previous discussion, we assume that $y_i' < y_i$. If the difference in quality is significant then the two scores will correspond to different EDTs, namely, $\delta_i^j \leq y_i' < \delta_i^k \leq y_i$. This translates to the following requirement:

> Given a significant difference in quality, i.e, $\delta_i^j \leq y_i' < \delta_i^k \leq y_i$, and the same a-priori confidence *Pr(A)*, it must be the case that $Pr(A|y_i) > Pr(A|y_i')$.

If the above requirement is met for our fusion strategy, then we say that our strategy is robust against changes in quality. Thus, considering Equation 19, we see that satisfying $Pr(A|y_i) > Pr(A|y_i')$ requires the following condition to hold:

$$\frac{TAR(y_i)}{FAR(y_i)} > \frac{TAR(y_i')}{FAR(y_i')}. \tag{21}$$

Since we have $\delta_i^j \leq y_i' < \delta_i^k \leq y_i$, Equation 21 requires that the TAR to FAR ratio at δ_i^k must be greater than the TAR to FAR ratio at δ_i^j. In fact, we can easily enforce the above condition in our error-rate-based fusion strategy by explicitly requiring that one should select the EDTs, $\{\delta_i^1, \ldots, \delta_i^m\}$, such that the following condition holds:

$$\frac{TAR(\delta_i^m)}{FAR(\delta_i^m)} > \cdots > \frac{TAR(\delta_i^1)}{FAR(\delta_i^1)}. \tag{22}$$

Considering Figure 8 and Figure 10 together, we note that for typical expert values the TAR to FAR ratio increases as y_i increases and vice versa. A good value for the EDT δ_i^1 will then commonly be the point of equal error rate.

Also note that if the quality of a sample is so poor that $y_i' < \delta_i^1$ then the sample is rejected. This is because $FAR(y_i')$ and $FRR(y_i')$ will evaluate to *1* and *0* respectively, as defined in Equation 14 and Equation 15.

5.4 SUMMARY

Combining scores from multiple biometric experts is known as sensor fusion. A common challenge in this field is that the results from evaluating different biometric characteristic are usually incompatible, as they have different score ranges as well as different probability distributions. Error-rate-based fusion is presented as a novel fusion technique that transforms individual scores from different biometric systems into objective evidences and combine them using Bayesian inference.

Error-rate-based fusion uses the false acceptance and false rejection rates of the biometrics systems to generate a confidence value that represents the probability that a principal has been correctly identified. For precise evaluation of the error rates a non-parametric approach is taken, where the false acceptance and false rejection rates are modelled as step functions.

The quality of the acquired biometric samples vary due to poor environmental conditions or the relative position of the user with regard to a remote biometric sensor. The presented error-rate-based fusion technique is shown to be robust to this change in quality under a set of reasonable assumptions.

Part IV

IMPLEMENTATION

6

PROTOTYPE

In this chapter we present the algorithms and techniques used in the implementation of our camera-based persistent authentication prototype. We start by looking at the representation of the target's state and the integration with authentication and authorisation zones. We then discuss the implementation of the tracker and the use of filtering and flow techniques to aid in person tracking. We continue with an overview of remote biometrics and the implementation of the chosen biometric characteristics, namely, facial recognition and appearance analysis. Finally we end with the implementation of the persistent authentication model and a discussion of the applications for location-based services and the challenges in multi-camera setups.

In order for a surveillance system to be useful in an interactive environment, the system needs to work in real time with data from multiple cameras. Therefore, the implementation must be extremely efficient. This implies a strong focus on parallelisation and distribution to integrated hardware. As a result, the most important design goal in the persistent authentication prototype is to create a real time, multi-camera surveillance system. We approach this goal by presenting a fast and efficient tracking methodology that builds upon well structured and fully developed computer vision algorithms for background modelling and foreground detection. We use error-rate-based fusion to combine the output of biometric experts to increase the confidence of the tracker. The persistent authentication model provides input to a location-based service, represented by a context-aware access control system, which enables location-based service provision, directly at the principal's position.

At each time instance, the persistent authentication prototype receives information about the state of the target principal and passes it in parallel to the authentication, tracking, the detection components. The tracker estimates the motion of the target based on its previous state and outputs a single hypothesis. The detection and recognition components maintain a model of the target and constantly analyse the output of the tracker and the state of the principal to detect and recog-

nise any appearances represented in the target model. This information is used by the fusion component, which also output a hypothesis about the location of the target. The hypotheses are passed to the persistent authentication component which combine them into a single score that reflects the system's confidence in the identity of the principal. This confidence score determines the provision of location-based services at designated authorisation zones in the environment. Figure 11 shows an overview of the components in the persistent authentication prototype.

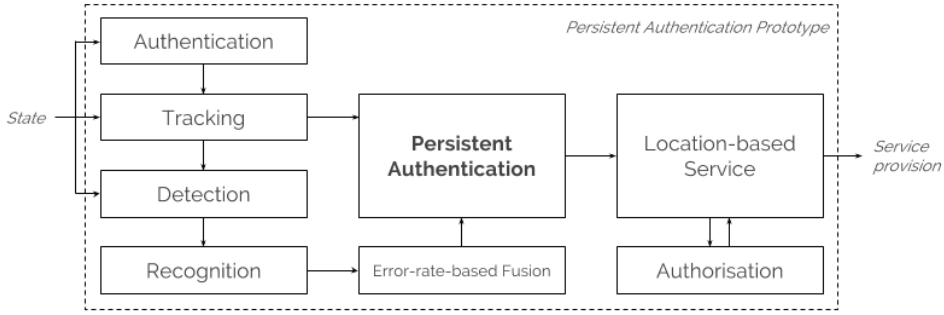


Figure 11.: Overview of the components in the persistent authentication prototype.

In the following we go into details with each of the components in the persistent authentication prototype. The implementation touches upon many areas in computer vision. Computer vision is a young field, but has seen tremendous progress in recent years. One of the driving forces behind this progress is the availability of high quality frameworks and libraries that enable fast and efficient processing of state of the art algorithms. In the persistent authentication prototype we make extensive use of the Open Source Computer Vision Library (OpenCV). OpenCV is an open-source BSD-licensed library that includes a comprehensive set of both classic and state of the art computer vision and machine learning algorithms. OpenCV is designed for computational efficiency, with a strong focus on real-time applications. It is written in optimised C/C++ and can take advantage of hardware acceleration and multi-core processing.

6.1 STATE

The state is the representation of the target principal. The state is the result of a feature extraction process on every frame of the video data as captured by the cameras in the environment. At any time instance, the state is defined by a bounding box detailing the position of the target. The bounding box is parameterised by its location and scale. The spatial similarity between two bounding boxes are measured using overlap. The state only concerns location and scale, other parameters such as in-plane rotation are not considered. Finally, the lack of a bounding box for a target indicates that tracking is not currently possible for that target, for instance because the tracked principal is not visible.

In the implementation we use background modelling to segment moving regions in the video. We assigning one label to the principals and another label to their surroundings and perform contour analysis to help draw the bounding boxes. We use a Gaussian Mixture Model algorithm based on work by Zivkovic et al. [138] [139], which automatically selects the appropriate number of components per pixel and thus are able to fully adapt to the observed scene. The implementation uses a non-parametric adaptive density estimation method, which reduces processing time and improves the segmentation.

An important consideration in the Gaussian Mixture Model algorithm is the approach for determining learning rates and initialisation parameters. The learning rate controls the adaptation and convergence speed of the model and introduces a compromise between being fast enough to adapt to changes and slow enough to store a useful temporal history. The optimal initialisation parameters and value for the learning rate depends on the given scene. In the original publication [87] [88] the learning rate is set to some fixed small value, which allows the mixture model to adapt to the gradual changes and prevents it from including foreground objects into the background. In our implementation we follow this approach by hand-tuning the learning rates and initialisation parameters in order to produce the desired subtraction result. The need to tune these parameters makes it difficult and time consuming to change setups, but gives a great deal of control over the end result. Alternatively, the learning rate and other parameters of the mixture model can be adjusted automatically, for instance using particle swarm optimisation [140] or spatio-temporal voting schemes with random samples [141], which are useful additions for more dynamic scenes.

In a comparative study Parks et al. [98], examine how a number of pre- and post-processing techniques can improve identification of blobs in the foreground mask. The authors conclude that noise removal, morphological closing, and area testing significantly improves the performance. Noise removal is beneficial as camera noise and limitations of the background model often introduce small specks of noise in the foreground mask. These errors can be removed by applying a noise filtering algorithm to the foreground mask.

The authors specify ρ as the number of foreground pixels amongst 8-connected neighbours, w as the width of a square kernel and a as the number of foreground pixels. We apply a density-based noise removal method to the foreground masks which discards a foreground pixel if it has less than 7 foreground pixels amongst its 8-connected neighbours. We then apply morphological closing to fill internal holes and small gaps in the blobs. We use a square 3–by–3 kernel, which slides through the image in two iterations. Finally, area thresholding is used to remove remaining blobs that are smaller than 30 pixels in size.

After applying the morphological operations we identify the contours in the image using a topological structural analysis of the foreground mask by border following [142]. Each contour is stored as a vector of points and enveloped by a bounding box. This bounding box describes the state of the target and is provided to the authentication, tracking and detection components.

6.2 AUTHENTICATION AND AUTHORISATION

The authentication and authorisation components define the zones placed in the environment as described in section 4.1. Tracking is initialised when a principal successfully authenticates in the authentication zone and service is provided to principals in authorisation zones, given they have sufficient clearance.

The zones are defined as a vector of points corresponding to the authentication and authorisation areas in the environment. In the implementation we abstract this definition by introducing the assumption that these areas can only physically hold a single principal at a time. This eases the process of linking the authentications and authorisations to the correct principal. The assumption requires a very strict security policy, but we found that this corresponded well with the security policies in place in the premises we investigated. Later, in chapter 8 we present

a review of the security policies in one of the investigated premises and how different layouts and rules for the accompanying authentication and authorisation zones affect performance.

For premises with less restrictive security, the link between the principals and the corresponding authentications and authorisations may be solved with additional sensors. Adding an additional sensor domain and fusing the data, gives the system another dimension to base its decisions on. Infrared cameras can be used for depth information and pressure sensitive floors provide very different operating characteristics to regular cameras. Integrating these sensor types allows the system to more reliably link the principal and authentication/authorisation which increases the robustness.

6.3 TRACKER

The tracker propagates the state of the target principal over time by estimating frame-to-frame motion. The tracker recursively use state information to form a trajectory for each of the tracked principals. The tracker initialise a track for every incoming principal and terminates the trajectories associated with disappearing targets. Initialisation always occurs at an authentication zone, whereas the trajectories must be terminated when a target leaves the field of view of the camera, or when tracking degrades under a predefined performance level. The tracker is an exploratory and error-prone component, and therefore it is expected that, without correction, tracking performance will degrade over time to the point of tracking failure.

The persistent authentication prototype relies on an efficient tracking methodology that adapts to changing conditions and inaccuracy in the sensor data. The tracking methodology is built on a combination of frame-to-frame tracking and tracking-by-detection. In frame-to-frame tracking it is assumed that the principal moves with a smooth trajectory. This approach can adapt to changes in the appearance of the principal, however, it will typically fail if the principal gets fully occluded or disappears. In contrast, tracking-by-detection assumes that a model of the principal is known in advance. This approach is resilient to occlusions and disappearances, but may fail in cases with unexpected appearances or cluttered backgrounds.

In our system we divide tracking into two categories with different time scales. First, we have *motion tracking* which is considered on a frame-to-frame basis while the principal is visible and easily identified. If the principal is momentarily occluded or there is temporary noise in the data, motion tracking relies on the last know position of the principal and the associated trajectory and velocity of the principal. If the tracking performances degrades noticeably or if the principal is absent for a longer period of time we suspend tracking and consider *restorative tracking*. Here, we require a re-authentication of the principal, either from a discrete authentication mechanism or a biometric expert. We then restore the tracking session and re-associate it with the target principal. Figure 12 gives an overview of the tracking component.

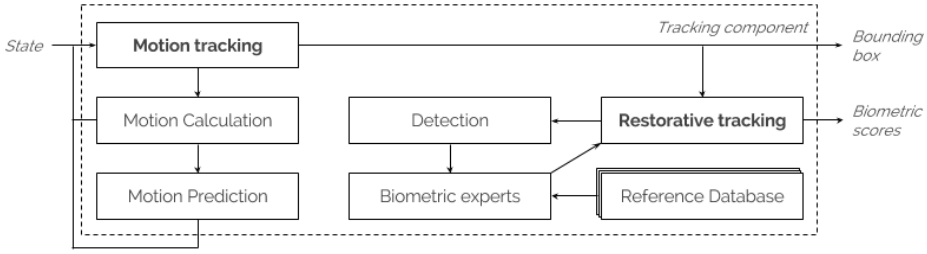


Figure 12.: Overview of the tracking component.

6.3.1 *Motion Tracking*

In motion tracking the target is represented by a bounding box on which a sparse motion field is calculated. We compute region correspondences based on the shape, location and motion features of the target. This gives a strong correlation when used on a frame-to-frame basis and since the process is not computationally intensive it is suitable for real time applications.

The motion fields for each principal is estimated by calculating the optical flow. Optical flow is the pattern of apparent motion of image objects between two consecutive frames, caused by the movement of the object. We use a pyramidal Lucas-Kanade approach [101] to calculate the motion points. These points form a 2D vector field where each vector is a displacement that shows the movement

of the point from the first frame to the second. In the implementation we integrate a sparse iterative version of the Lucas-Kanade optical flow tracker [143] which supports parallelisation for efficient calculation. In addition, this implementation of the Lucas-Kanade tracker selects appropriate features within the bounding box and applies a backwards-check of the feature points to reduce the number of points that are wrongfully estimated or disappears between images.

The tracker receives the bounding box b_t of the target principal and a pair of images I_t, I_{t+1}. A set of points is initialised within the bounding box b_t and the points are tracked by the Lucas-Kanade tracker from I_t to I_{t+1}. The tracker then outputs the resulting bounding box b_{t+1}. The bounding box motion is parameterised by the horizontal and vertical displacement and the change in scale. All three parameters are estimated independently using the median. Figure 13 shows an example of three principals, each described by a bounding box, where the arrow in the centre of each bounding box illustrates the median flow.

Figure 13.: Three principals described by bounding boxes. The arrow indicates the median flow.

6.3.2 *Tracking Failures*

The frame-to-frame tracker can fail due to temporary occlusions or adverse sensor noise. To detect these failures we let δ_i denote the displacement of a point from I_t to I_{t+1} and δ_m denote the median displacement of all points. We declare a tracking failure if the median absolute deviation is larger than a threshold, as given by $\text{median}(|\delta_i - \delta_m|) > \Delta$, where a typical value of Δ is 15 pixels.

Using this heuristic the tracker is able to identify most failures caused by fast motion or sudden occlusions of the target. In these cases, the displacement of the

individual points become scattered in the image, causing the median absolute deviation to increase rapidly.

If a failure is detected, the tracker does not return a bounding box. Instead the tracker relies on motion consistency to follow the target.

6.3.3 *Trajectory Hypothesis*

We create a trajectory hypothesis that predicts the location of the target over time. We let a Kalman filter process the position of each principal in the previous frames, to estimate their position in the current frame. The filter's estimate gives a prediction of the motion of the principals.

We use the Kalman filter algorithm presented by Welch et al. [106]. In our case, we use a Kalman filter with 4 dynamic parameters and 2 measurement parameters and no control. The algorithm assumes that the state vector $\mathbf{x}_t$, at time t is evolved from the previous state *(t–1)* using the state transition matrix $\mathbf{F}_t$ that relates the current time step to the previous one and the measurement model $\mathbf{H}_t$ that relates the state to the measurement:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{w}_t \\ \mathbf{z}_t &= \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t \end{aligned} \tag{23}$$

With the state vector $\mathbf{x_t}$ of the form:

$$\mathbf{x} = [x, y, vx, vy]^t \tag{24}$$

Here $\mathbf{x}$ is a model vector, $\mathbf{z}$ is a measurement vector, $\mathbf{w}$ is the model noise drawn from the distribution identified by the process noise covariance matrix $\mathbf{Q}$ and $\mathbf{v}$ is the measurement noise drawn from the distribution identified by the measurement covariance matrix $\mathbf{R}$. Initialisation consists of constructing a model for the transition matrix $\mathbf{F}$, the measurement model matrix $\mathbf{H}$ and the noise matrices.

It is assumed that the parameters are independent and that there is no need for transformation to convert measurements into model data. Therefore, the $\mathbf{H}$ matrix is initialised as an identity matrix and the $\mathbf{R}$ and $\mathbf{Q}$ matrices are initialised as one-valued diagonals with pre-defined values.

The **F** matrix represents the expected transformation of a given state and shows how the parameters change with each time step:

$$\mathbf{F} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{25}$$

The next time step can then be evolved from the previous one, using the discrete approximation of the equations for an object moving with a constant velocity model:

$$\begin{aligned} x_{t+1} &= x_t + vx_t \\ y_{t+1} &= y_t + vy_t \\ vx_{t+1} &= vx_t \\ vy_{t+1} &= vy_t \end{aligned} \tag{26}$$

An example showing the Kalman filter predicting the position of a principal through an occlusion is shown in Figure 14. The small circle shows the prediction of the filter, which correctly follows the target (large circle).

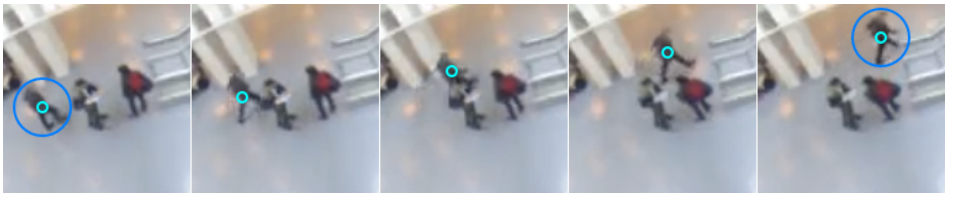


Figure 14.: The Kalman filter prediction (small circle) predicts the path of a tracked principal (large circle) through an occlusion.

To re-associate the tracking sessions, we need to solve the matching problem, i.e., associate the prediction of the filter with the target principal. The simplest solution to the matching problem is to use a nearest neighbour approach. However, when tracking multiple principals in a crowded scene, there is no guarantee that this approach finds the correct principal. Instead we take advantage of the fact that the motion of principals rarely change drastically in small time windows.

Thus, we compare the previous state with the current state and the previous motion field with the current velocity of the principal. Assuming motion consistency, this gives a good indication of the target principal.

Figure 15 shows an exemplified position-time graph of the x-coordinate of the tracked principal in Figure 14. The graph shows how the tracker begins to drift from the ground truth due to the occluding principals. The track is terminated and for a number of frames the principal is obscured, thus no bounding box is outputted. The predictions from the Kalman filter are used as search regions, within which the tracker tries to relocate the target principal in each of the following frames. In the example on Figure 14, the tracker is successful in re-associating the tracking session with the target principal after the occlusion. This is illustrated in Figure 15 as the point where the motion estimate reaches the ground truth.

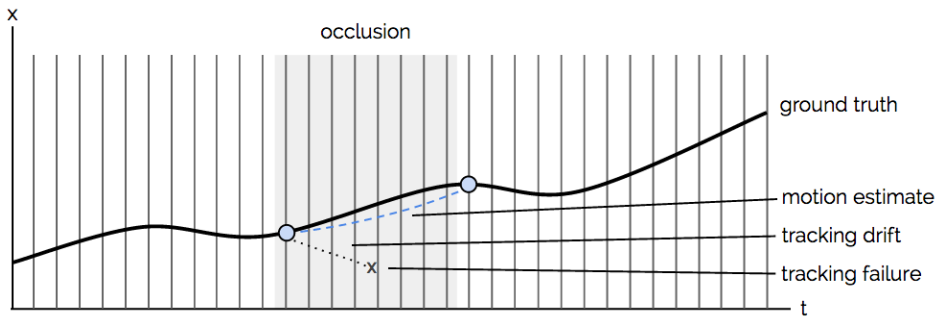


Figure 15.: Position-time graph showing an example of tracking failure and re-association based on motion estimation.

The result of the re-association is forwarded to the persistent authentication component, which updates the confidence in the identity of the principal. If this update reduces the confidence below a threshold, tracking is suspended. The tracking mode is then changed to *restorative tracking* and a re-authentication of the principal is required to restore the tracking session.

6.4 DETECTION, RECOGNITION AND FUSION

The current state of the principal and the output of the tracker is used by the detection and recognition components. These components maintain a model of the target and constantly analyse the output of the tracker and the state of the principal to detect and recognise any appearances represented by the target model. A match is used to increase the confidence of the tracker or to restore a suspended session. The components are populated with prior information about the principal and updated with online information. The output is fused using error-rate-based fusion to increase the robustness of the evaluation.

6.4.1 *Remote Biometrics*

Biometrics are a multidimensional problem with the aim of understanding the uniqueness of humans to facilitate recognition and verification of their identity. Remote biometrics operate at a distance and require no interaction from the users, thus ensuring a calm authentication process.

In the persistent authentication prototype we utilise remote biometrics to perform calm authentication of users. We consider two user-centric remote biometrics, namely, facial recognition and appearance analysis. We let the biometric experts perform continuous authentication by processing samples of the biometric modalities as they become available. We fuse the scores using error-rate-based fusion and adjust the confidence in identify given the fused score y_i.

To detect faces we use the object detection framework presented by Viola and Jones [10] [11]. The detection is eased by leveraging the tracking information as an initial search region for the algorithm. This significantly speeds up the detection of faces in the persistent authentication prototype and allows faces to be detected in real time with limited processing overhead.

In the following we discuss the implementation of facial recognition and appearance analysis.

6.4.1.1 *Facial Recognition*

For facial recognition we use a linear subspace technique to project high dimensional data into a lower dimensional subspace by linearly combining features.

Principal Component Analysis (PCA) [144] and *Linear Discriminant Analysis* (LDA) [145] are well established linear subspace techniques and are considered some of the most robust methods for face recognition [146]. In the following we present PCA and LDA and describe their differences.

Consider a set of N facial images $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$ with values in an n-dimensional image space. A linear transformation maps this n-dimensional image space into a lower m-dimensional feature space $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N$ such that $\mathbf{y}_k$ represents $\mathbf{x}_k$ by introducing a transformation vector W such that:

$$\mathbf{y}_k = W^T \mathbf{x}_k \qquad\qquad k = 1, 2, ..., N \tag{27}$$

For the transformation to accurately represent the original data, it is important to retain the highest possible variation, thus the objective is to find a subspace in which the variance is maximised. Let the total scatter matrix S_T be defined as:

$$S_T = \sum_{k=1}^{N} (\mathbf{x}_k - \mu)(\mathbf{x}_k - \mu)^T \tag{28}$$

Where μ is the mean of all the images. After applying the linear transformation W^T, the scatter of the transformed feature vectors $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N$ is $W^T S_T W$. In PCA, the projection W_{opt} is chosen to maximise the determinant of the total scatter matrix of the projected samples, such that:

$$W_{opt} = \arg\max_{w} (W^T S_T W) \tag{29}$$

The output is a set of n-dimensional eigenvectors $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m$ corresponding to the m largest eigenvalues, which account for the most variance in the training set. Since these eigenvectors have the same dimension as the original images, they are referred to as Eigenfaces [144].

In PCA, classification can be performed in this reduced feature space, for instance using a nearest neighbour classifier. However, a drawback of this approach is, that much of the variation we seek to maximise is caused by illumination changes [147], thus with images of faces under changing illumination the projected feature space will contain variation due to lighting and not necessarily due to class separability. Consequently, the points in the projected space will not be well clustered. A better approach is to use Linear Discriminant Analysis, where classification is performed by selecting W in such a way that the ratio of the

between-class scatter S_B and the within-class scatter S_W is maximised. With the between-class scatter matrix defined as:

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T \tag{30}$$

and the within-class scatter matrix defined as:

$$S_W = \sum_{i=1}^{c} \sum_{\mathbf{x}_k \in X_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^T \tag{31}$$

where μ_i is the mean image of class X_i, where $X_i = 1,2,...,c$ and N_i is the number of samples in class X_i. A projection, W_{opt} is then found, that maximises the class separability criterion:

$$W_{opt} = \arg\max_W = \frac{|W^T S_B W|}{|W^T S_W W|} \tag{32}$$

The benefits of the class specific linear projection of LDA in comparison to PCA is traditionally illustrated using the Iris flower dataset, a multivariate dataset introduced as an example of discriminant analysis. The data quantify the morphologic variation of Iris flowers of three related species *"all from the same pasture, and picked on the same day and measured at the same time by the same person with the same apparatus"* [148].

The dataset consists of 50 samples from each of three species of Iris setosa, Iris virginica and Iris versicolor. Four features were measured from each sample: the length and the width of the sepals and petals.

In the implementation we use the algorithms detailed by Duda et al. [149] for Principal Component Analysis and the algorithm presented by Belhumeur et al. [146] for Linear Discriminant Analysis.

Principal Component Analysis applied to the data identifies the combination of attributes that account for the most variance in the data, whereas Linear Discriminant Analysis try to identify attributes that account for the most variance between classes. Figure 16 plots the two first principal components and the result of the Linear Discriminant Analysis. As seen on the figure Linear Discriminant Analysis allows for better class separation on the iris dataset.

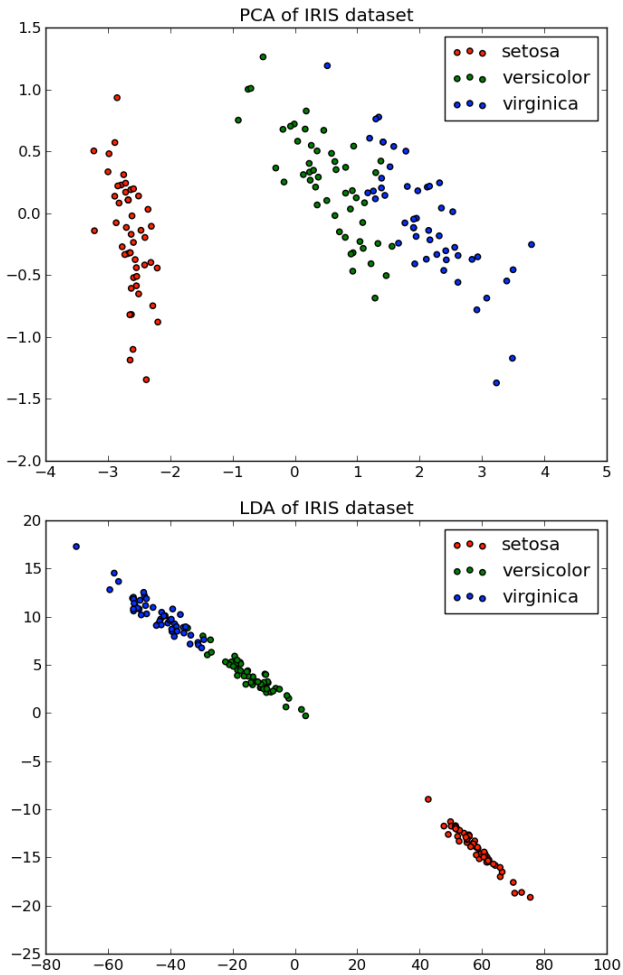


Figure 16.: Principal Component Analysis (top) and Linear Discriminant Analysis (bottom) on the Iris dataset

When used for facial recognition Linear Discriminant Analysis is known as *fisherfaces*, after the original author of the algorithm. The performance of fisherfaces is dependent on a well founded training set. In the persistent authentication prototype we construct this training set from a series of enrolment images, which we augment with high quality faces captured during operations. In the absence of enrolment images our prototype can operate using only the captured faces, though at a loss of accuracy.

6.4.1.2 *Appearance Analysis*

For appearance analysis we use colour profiles of the principals, calculated using histogram comparison. Colour histograms are widely used for content-based image retrieval [150] as they are fast to compute, and despite their simplicity, have attractive properties. Since they contain no spatial information they are largely invariant to rotation and translation of objects in the image. Additionally, colour histograms are less affected by partial occlusions and changes in camera viewpoint than facial recognition [151] [152].

The appearance of a person is considered only as auxiliary information in biometrics, as it is not by itself sufficient to establish the identity of a person because it generally is indistinct, unreliable, and can be easily spoofed. Appearance is known as a soft biometric, which is defined as characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals [153].

Although soft biometrics are not as permanent or reliable as traditional biometric characteristics, such as fingerprints and facial recognition, they do provide important information about the identity of the user, that can be leverage in combination with other biometrics to provide higher accuracy in establishing the principals identity.

Colour histograms used for appearance analysis are typically represented in the RGB or the HSV colorspace, where the latter representation is more robust with respect to illumination variability [154].

The difference between two histograms h_1, h_2 can be expressed, for instance by the chi-squared distance, as follows:

$$\chi^2(h_1,h_2) = \frac{1}{2}\sum_k \frac{(h_{1k}-h_{2k})^2}{h_{1k}+h_{2k}} \tag{33}$$

The histograms are calculated on the same regions as the facial recognition expert. The search regions provided by the tracker are considered as bounding boxes for the histograms. The histograms are calculated using the foreground mask, to filter out any background information, which increases the robustness of detection and reduces the processing time.

We combine the calculated distances with the results from the facial recognition using the error-rate-based fusion scheme described in chapter 5. The fusion process yields a more robust result than provided by each of the individual systems. The fused biometric score y_i is used to update the confidence in the identity of the principal $Pr(a|y_i)$ in the persistent authentication component.

6.5 PERSISTENT AUTHENTICATION COMPONENT

The persistent authentication component combine the responses from the tracker and the fusion component into a single response and outputs the final hypothesis about the target's state. The updated score reflects the system's confidence in the identity of the principal and determines the provision of location-based services.

The confidence in the identity of the principal is calculated using Equation 7. The confidence after receiving the authentication score s_i is expressed as the conditional probability $Pr(a|s_i)$. Likewise the confidence in identity after receiving a biometric similarity score from the error-rate-based fusion component, y_i, is expressed as $Pr(a|y_i)$. The noise e_i that will invariably occur in some of the measurements results in an update of the confidence in a similar fashion $Pr(a|e_i)$. This approach is illustrated in Equation 34:

$$Pr(a|s_i) = \frac{Pr(s_i|a)Pr(a)}{Pr(s_i)} \tag{34}$$

Ultimately, the updated score reflects the system's confidence in the identity of the principal and determines the provision of location-based services.

When tracking is lost, due to an occlusion or noise in the sensor data, the system's confidence is lowered accordingly. We take an approach similar to Luber et al. [155] and Mucientes et al. [156]. We let the probability of correctly re-identifying a principal be described by an exponential function. This simulates the decay in the probability of detecting a principal that has not been matched for several consecutive frames. More formally, let $t - t_0$ be the number of consecutive frames that principal a has not been observed, the probability of the event e_i given the principal a is then defined as:

$$Pr(e_i|a) = exp(-\frac{t-t_0}{\lambda_i}) \tag{35}$$

where λ is the speed of the decay process. This exponential distribution ensures a gracefully degradation of the tracking confidence, as the confidence decreases concurrently with the likelihood that a principal will be correctly identified as time passes.

In many cases the tracker fails when the paths of multiple principals intertwine or when principals move too close together for the camera-based sensors to differentiate them. In these cases the persistent authentication component will revert to a failsafe state, setting $Pr(a)$ to a fraction of the likely candidates, e.g., $Pr(a)$ would be set to 0.5 if two principals are equally likely. This retains some of the authentication confidence, and may increase the usability of the system, as the principals can still access low-priority areas in the building, where only a modest confidence score is required.

6.6 LOCATION-BASED SERVICE

In real world applications, location-based services need to cover large areas that may be composed of multiple cameras with overlapping and/or disjointed field of views. In addition, these location-based services often include multiple authentication and authorisation zones spread out in the environment. The principals may appear in multiple cameras and their appearance in one camera might be very different from their appearance in the next, for instance due to illumination and pose differences or different camera properties. In addition, the occupants may require any number of services in many locations. In conclusion, providing location-based services is a challenging task that requires careful consideration.

In the following we present some of the methods we use to make location-based services possible in extensive environments and highlight an interesting application for location-based services that is related to the behavioural analysis of principals in the environment.

6.6.1 *Multi-camera Systems*

Multi-camera systems rely on the correspondences between a number of cameras distributed in the environment. The cameras may be positioned such that they monitor the same scene from different angles. For person tracking, this generally gives more robust tracking results, especially in cases with persistent occlusion between principals as the different viewing angles allows better segmentation of the principals. However, it is not always possible to have overlapping camera views due to limited resources or large areas of interest. Non-overlapping multi-camera systems have to to deal with sparse object observations. Therefore additional assumptions have to be made about the behaviour of the principals in order to correlate observations from one camera to the next. Conversely, for overlapping camera systems, calculating the correspondence between observations from the same scene requires careful consideration.

In either case, the first step in a multi-camera system is to identify the actual area covered by the cameras in the world state and find a projection that maps this into the model. In the analysis of digital images this problem is not always trivial, as camera lenses introduce distortion in the images. Lens distortion occurs in all cameras and one of the most prevalent forms of distortion is known as barrel distortion. It results from the lens having a slightly higher magnification in the centre of the image than in the periphery, which causes lines in the image to appear warped around the centre. Barrel distortion is particular noticeable in cameras with wide-angle lenses, which is clearly seen on the left in Figure 17.

Barrel distortion is primarily radial, and to correct barrel distortion one method is to find the distortion parameter of the camera and use it to correct the distortion. Li et al. [157] found that a relatively simple one parameter model can be used to account for most of the distortion. Bailey et al. [158] propose a solution to this model based on fitting parabolas to the distorted lines in the image and using

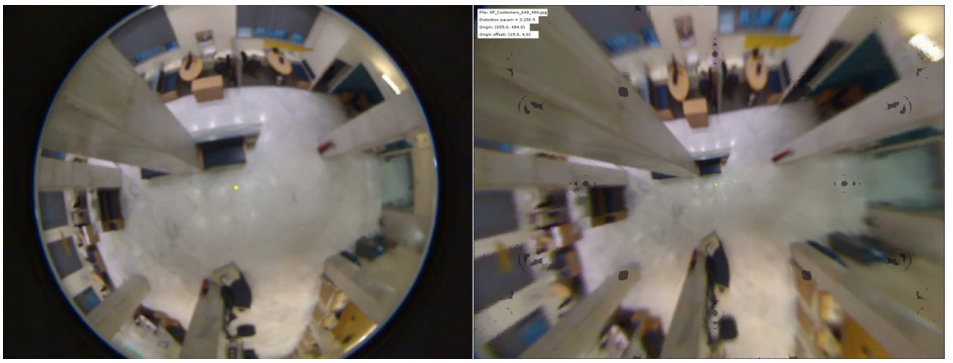


Figure 17.: Distorted image (left) and corrected image (right)

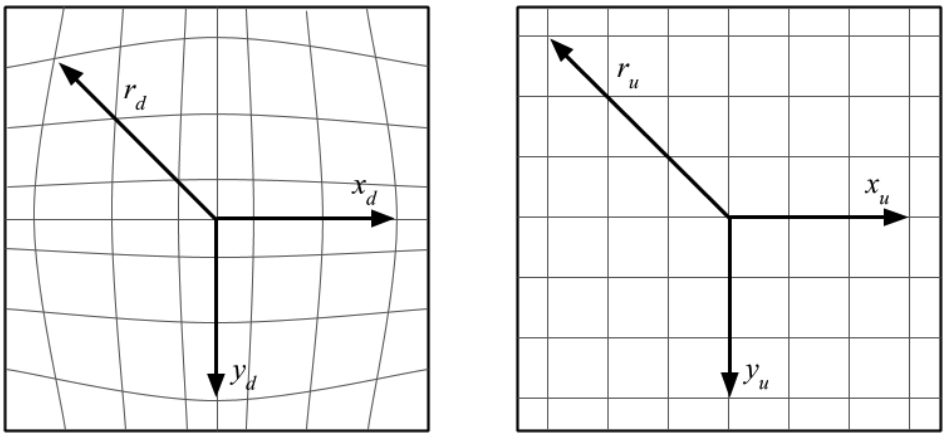


Figure 18.: Illustration of the barrel distortion model

the curvature to estimate the radial distortion component. The barrel distortion model the authors use is:

$$r_u = r_d(1 + \kappa r_d^2) \tag{36}$$

where r_u and r_d are the distance from the centre in the undistorted and distorted images respectively and κ is the distortion parameter, which is specific to the camera lens. Figure 18 shows an illustration of the barrel distortion model.

κ is determined for each of the fitted parabolas and a weighted average is calculated. The barrel distortion can then be corrected by calculating the coordinates in the undistorted image as a function of those of the distorted image. However, in most cases we are interested in determining the inverse, i.e., we want to use the coordinates in the undistorted image to select which pixel to display from the distorted image. Therefore the barrel equation must be of the form:

$$r_d = F(\kappa, r_u) \tag{37}$$

Gribbo et al. [159] present a real-time implementation that solves this formulation of the barrel distortion equation. Their approach also uses bilinear interpolation to improve the quality of the corrected image. The result from correcting barrel distortion is shown in Figure 17.

With the distortion corrected, the next step in a multi-camera system is to identify the correspondences between the cameras. In computer vision, many automated methods exists to this problem and the survey by Szeliski et al. [160] present an excellent overview of the most popular ones. On the other hand, manual adjustment of the correspondences provides a simple and accurate alternative to the automated methods, however it may be quite time consuming for a larger number of cameras.

In the installations we have conducted with the persistent authentication prototype the number of cameras have been relatively low, in part due to the large area covered by the wide-angle lenses, and in part to minimise the disruption caused by the installation. As a result we have relied on manual adjustment of the camera correspondences. Figure 19 shows how the corrected image from Figure 17 have been fitted to a model of the considered environment. The blue line represents the floor plan of the environment. The purple line indicates an input mask applied to avoid certain regions, e.g., the windows to the street. The green line

shows the camera view after taking into account the topology of the building and the interposing objects.

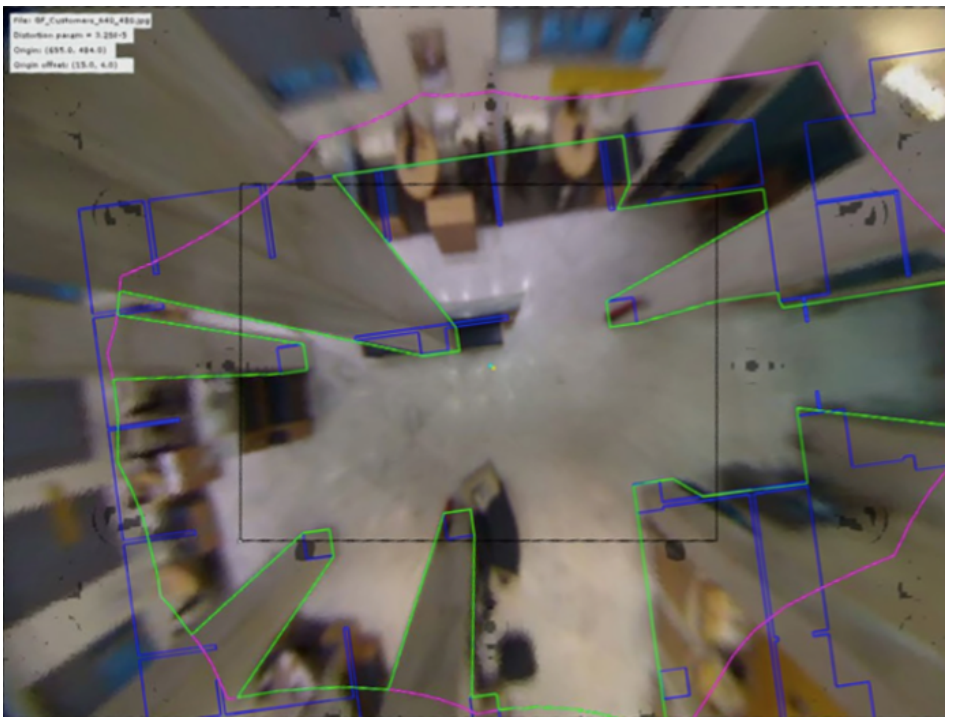


Figure 19.: Corrected image fitted to the environment model

6.6.2 *Behavioural Analysis*

For a location-based service that spans multiple cameras, an interesting application presents itself, namely the behavioural analysis of the principals in the environment.

In the persistent authentication prototype we utilise the calculated optical flow fields to form spatio-temporal patterns in the movement of the principals. Spatial coherence describes the correlation between signals at different points in space, whereas temporal coherence describes the correlation or predictable relationship

between signals observed at different moments in time. By stacking multiple consecutive frames from an image sequence on top of each other we obtain a spatio-temporal volume that has two spatial dimensions and a temporal dimension.

An example of such aggregated motion vectors, calculated using optical flow, is shown on Figure 20. By aggregating the flow, the dominant patterns of movement in the building are identified. These patterns, in combination with the spatial analysis of the building, can be used to detect and predict behaviour in the building, as we later show in section 8.1.

6.7 SUMMARY

An overview of the algorithms and techniques used in the implementation of our camera-based persistent authentication prototype have been presented. The prototype involves three main steps: detection of people in the scene, tracking of these people, and evaluation of the tracking results with regards to the movement patterns and the behaviour of the occupants.

We use motion-tracking for immediate frame-to-frame tracking of principals. In cases with occlusions or noisy measurements motion tracking relies on trajectory hypotheses and motion flow to re-identify principals. The probability of correctly identifying a principal is modelled with an exponential function to represent the decreasing likelihood of identifying the correct principal over time. Finally, if a principal is occluded for a longer period of time we consider restorative tracking. Here, the authentication session is suspended and requires a re-authentication before restoring the session.

Facial recognition and appearance analysis are integrated in the persistent authentication prototype as remote biometric experts that operate at a distance and require no interaction from the users. The experts perform continuous authentication by processing samples of the biometric modalities as they become available. The results are fused with error-rate-based fusion to increase the robustness of the evaluation.

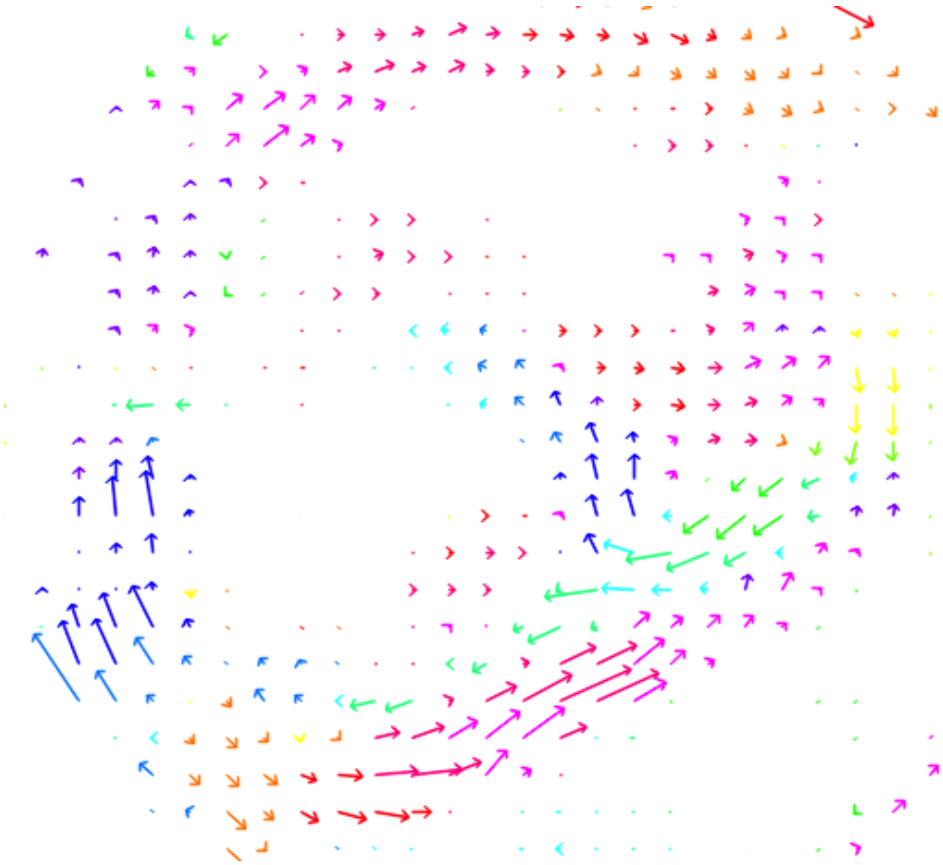


Figure 20.: Motion vectors calculated for each pixel in the image, aggregated in a 20x20 grid, with magnitude (length) and the direction (colour).

Part V

EVALUATION

7

EVALUATION FRAMEWORK

In this chapter we present the framework and methods used in the evaluation of the camera-based persistent authentication prototype. The evaluation consists of a combination of case studies, scenario-based validations and benchmarks on public-domain datasets that are commonly used in the literature.

Firstly, We assess the performance of biometric verification with the two chosen methods, Linear Discriminant Analysis and appearance analysis with colour histograms. We assess the performance of each biometric expert individually, and proceed with an evaluation of error-rate-based fusion of the biometric characteristics. The results are compared with the state of the art in score level fusion. This analysis is carried out on public-domain datasets to ensure the reproducibility of our results.

Secondly, we analyse the performance of the tracker using the following evaluation framework: the tracker is initialised in the first frame of a sequence and tracks the selected principal until a tracking failure occurs or the end of the sequence is reached. The generated trajectory is compared to the ground truth and the performance is measured as the number of frames the principal was successfully tracked. The analysis first evaluates motion tracking to establish the baseline performance of the tracker. Restorative tracking is then evaluated using continuous authentications provided by the biometric experts. The analysis is carried out on two public-domain dataset, the first used to assess motion tracking and the second to assess restorative tracking.

Finally, in order to ensure the convergence of the efforts deployed as part of the RIBS project, a test facility has been provided for the partners to evaluate their protection measures. Thus, we conclude the evaluation with our findings from installing the persistent authentication prototype in a live building. This part of the evaluation describes the data exploration techniques and tools developed to identify the contextual and behavioural patterns in the data and their application to intrusion detection. In addition, a series of scripted scenarios have

been defined which we use to perform scenario-based validation of the persistent authentication prototype.

For comparative purposes with regard to previous work, we evaluate the accuracy of the contextual awareness provided by different smart environments and the impact of using different sensing technologies. We employ the same metrics and test cases as presented by Kirschmeyer et al. [5] in their evaluation of the original *PAISE* prototype. The implementation of the prototype is evaluated on the basis of its persistence, robustness, and scalability in various conditions, including both expected use, usurpation and malicious behaviour as follow:

> **Persistence:** Addresses how well the system maintains the functional requirements of tracking, i.e., the system's ability to track principals and preserve authentication sessions.
>
> **Robustness:** Addresses the system's ability to resist malicious attempts of manipulation, such as usurping the identity of legitimate users or accessing services without proper authorisation.
>
> **Scalability:** Addresses the performance of the system to simultaneously authenticate and track a large number of principals in an extensive environment.

The performance of the persistent authentication system deployed in either a smart environment using closed-circuit television cameras (CCTV) or infrared time-of-flight cameras (TOF) are compared and the impact on the persistence, robustness and scalability noted. This part of the evaluation is presented in Appendix A.

7.1 DATASETS

To ensure the reproducibility of our results we evaluate the persistent authentication prototype on public-domain datasets. We consider four dataset, the IMM Face Database [161], the Biometric Scores Set (release 1) from the National Institute of Standards and Technology [162], the CAVIAR dataset from INRIA Labs at Grenoble in France, and the CAVIAR dataset from a shopping centre in Lisbon, Portugal [163]. The IMM Face Database and the Biometric Scores

Set is used to evaluate our error-rate-based fusion strategy with regard to the biometric modalities. The CAVIAR INRIA Labs dataset is used to evaluate the motion tracking capabilities of the persistent authentication system and finally, the Lisbon CAVIAR dataset is used to evaluate the performance of the persistent authentication prototype with continuous authentications provided by the remote biometrics.

7.1.1 *IMM Face Database*

The IMM Face Database consists of 240 still images of 40 different human faces, all without glasses. The images in the dataset are in high resolution format and have been resized prior to the tests to 128 x 96 pixels. The gender distribution is 7 females and 33 males. The pose, expression and illumination vary for each subject as follows:

1. Full frontal face, neutral expression, diffuse light.
2. Full frontal face, happy expression, diffuse light.
3. Face rotated approx. 30 degrees to the person's right, neutral expression, diffuse light.
4. Face rotated approx. 30 degrees to the person's left, neutral expression, diffuse light.
5. Full frontal face, neutral expression, spotlight added at the person's left side.
6. Full frontal face, arbitrary expression, diffuse light.

An example of the varying poses and illumination of a subject is shown in Figure 21.

7.1.2 *NIST BSSR1 Dataset*

The National Institute of Standards and Technology (NIST) Biometric Score Set (BSSR1) [162], is a public-domain dataset that is widely used in literature for

Figure 21.: IMM Face Database: The varying poses and illumination of a subject

benchmarking score level fusion strategies [39] [164] [165]. We use the NIST-face database which contains a set of similarity scores for each of two mid–2002 face recognition algorithms, labeled system *c* and system *g* respectively. Scores are provided on images from 3000 individuals. For each individual, the set contains one score resulting from a genuine comparison and a number of scores resulting from impostor comparisons as follows:

NIST-face Dataset
Number of subjects: 3000
Number of face systems: 2
Number of face images from which scores came: 2 per subject
*Number of scores: 2 * 3000 * 6000*
Number of similarity files: 12000

7.1.3 *CAVIAR INRIA Labs dataset*

The INRIA CAVIAR dataset contains a number of video sequences that include staged and scripted behaviour intended to test various forms of interactions. The videos are shot with a wide angle lens, and shows the entrance lobby of the IN-

RIA Labs at Grenoble, France. The resolution of the video is half-resolution PAL standard (384 x 288 pixels, 25 frames per second) compressed using MPEG2.

For each video sequence a hand-labelled frame-by-frame ground truth is provided, that details the location of each principal in the scene, and for some of the clips, additional information about their head, gaze direction and hand, feet and shoulder positions. The ground truth also contains bounding boxes for groups of principals. Figure 22 shows a typical frame from the dataset, where the individual principals have been annotated with boxes (yellow) and the group annotated with a bounding box (green).

The mapping between model and world state is computed using the ground plane homography information. Figure 23 shows the coordinate system and a set of reference points. Table 5 lists the corresponding positions.

Table 5.: INRIA Ground plane homography

Point	Position (x,y)	Position (cm)
1	(64,88)	(0,671.5)
2	(211,40)	(1116,670)
3	(349,184)	(1545,190)
4	(39,187)	(0,0)

7.1.4 *CAVIAR Lisbon dataset*

The second CAVIAR dataset is from a shopping centre in Lisbon, Portugal. This set focuses on person interaction, including people walking, meeting with others, conversing, and window shopping. The sequences show a corridor in the shopping centre with entrances to a number of shops. Each sequence is made up of two time-synchronised videos with two different view points. The first view point (a) shows a view along the length of the corridor and the second (b) shows a view across the corridor. The resolution of the videos are half-resolution PAL standard (384 x 288 pixels, 25 frames per second) and they are compressed using MPEG2.

Similar to the INRIA dataset, a hand-labelled frame-by-frame ground truth is provided for each video and the mapping between model and world state is

Figure 22.: Ground truth of individuals (yellow) and groups (green)

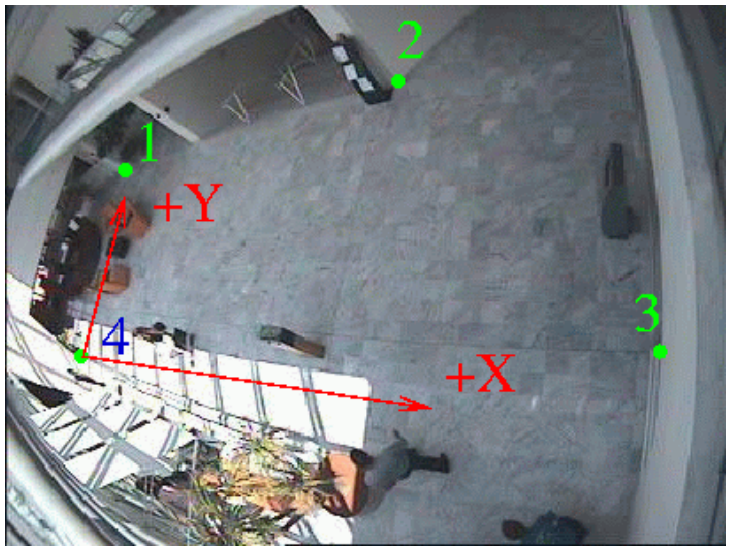


Figure 23.: INRIA Ground plane homography system and reference points.

computed using the ground plane homography. The homography coordinate system is shown for view (a) in Figure 24 and for view (b) in Figure 25, with the corresponding reference points listed in Table 6.

Table 6.: Lisbon Ground plane homography

Point	Position (x,y)	Position (cm)
1	(91,163)	(0,975)
2	(241,163)	(290,975)
3	(98,266)	(0,–110)
4	(322,265)	(290,–110)
5	(60,153)	(0,0)
6	(359,153)	(0,975)
7	(50,201)	(382,098)
8	(367,200)	(382,878)

7.2 EVALUATION OF BIOMETRIC FUSION

In the following we consider the performance of each of the two implemented biometric experts. We then fuse the scores using error-rate-based fusion and compared the results against the state of the art.

Score level fusion generally offers a good trade-off in terms of information content and ease in fusion. However, score level fusion is still a challenging task as the scores generated by the different experts can be either distance or similarity based and may follow different probability distributions. Additionally, the experts may provide quite different accuracies, which is accentuated when fusing hard and soft biometrics.

7.2.1 *Biometric Experts*

In the implementation of the persistent authentication prototype we use Linear Discriminant Analysis for face recognition and colour histograms for appearance analysis. Both methods operate on data that can be measured from a distance,

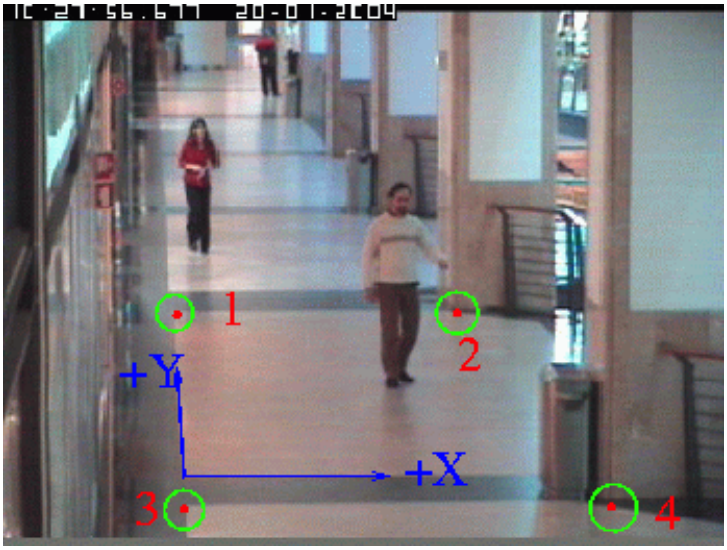


Figure 24.: Lisbon ground plane homography for view (a)

Figure 25.: Lisbon ground plane homography for view (b)

thus they are non-invasive and allows authentication to be performed continuously by sampling the modality recurrently.

We use the IMM Face Database to evaluate the performance of the biometric experts. We compare each image against every other image in the dataset. This generates a number of genuine scores, when the image pairs are from the same individual and conversely, a number of impostor scores when the image pairs are from two different individuals.

We calculate the error rates (ER) for each expert using different operating parameters. The error rates and the corresponding true acceptance rate (TAR) and false rejection rate (FAR) of the experts are shown in Table 7 for different thresholds.

Table 7.: Error rates of the biometric experts at increasing thresholds

Biometric Expert	**TAR**	**FAR**	**ER**
Face Recognition	89.1%	0.6%	5.75%
	95.8%	1.41%	2.82%
	99.6%	2.82%	1.61%
Apperance Analysis	70.77%	0.9%	15.07%
	83.5%	1.81%	9.17%
	100%	8.06%	4.03%

The results show that the LDA method described in subsection 6.4.1 is very accurate and robust to the varying poses, with the best error rate being 1.61%. The colour histograms have a comparatively higher error rate of 4.03%. The performance difference is to be expected, and shows that the more robust LDA method is capable of achieving high recognition rates, with a low false acceptance rate, when the size of the dataset is small.

As expected, we found that the majority of errors were due to the high variance between the images with either a rotated view or the added spotlight (images 3, 4 and 5 in Figure 21). As both pose and illumination affects the accuracy of the experts such images should generally be avoided for use as enrolments images in the system.

7.2.2 *Sum Rule Fusion*

We fuse the scores generated by the experts using error-rate-based fusion and compare the results to the *sum rule* combining scheme. In literature the sum rule is often used as the golden standard for comparison, as it provides a common ground for different score level fusion techniques. The motivation for using a common benchmark, instead of directly comparing the state of the art, arise as most fusion techniques are closed-source and difficult to implement. In contrast, datasets are readily available, and with the sum rule, enables the ranking of different fusion techniques across implementations based on their performance against the benchmark. Thus, we use the sum rule combining scheme to generate benchmark measurements that help asses the performance of our error-rate-based fusion technique.

The sum rule is a simple fusion operator and as noted by Norman et al. [37] is widely used for comparative purposes. The sum rule has been shown by Kittler et al. [40] [41] to outperform other simple fusion operators such as min, max, majority vote and product. Kittler and Alkoot [42] have later shown that for Gaussian distributions of estimation errors, the sum rule outperforms all other simple fusion operators. Fatukasi et al. [166] builds on the work of Kittler et al. and concludes that the sum rule should be used if either a high level of noise is present or when experts are highly correlated, which lends itself to the use of the sum rule for our comparison.

The sum rule assigns weights to each of the experts and fuse the data as shown in Equation 38. In the following evaluations we always consider an equal weighting.

$$y_{wsum} = \sum_{i=1}^{N} w_i y_i \tag{38}$$

In transformation-based score level fusion schemes, such as the sum rule, it is often necessary to normalise the input scores before the fusion process. Common approaches to score normalisation include: min-max normalisation, z-score normalisation, and tanh-estimator normalisation. He et al. [167] present an overview of these common normalisation techniques and compare the performance against a proposed method, namely reduction of high-scores effect (RHE) normalisation.

The authors found that multimodal biometric systems generally suffer more from low genuine scores, as opposed to high impostor scores. We also noted this effect in the discussion of the effect of quality on fusion in section 5.3. The situation occurs as the degrading quality of the biometric samples adversely affects the similarity scores. Thus, it is unlikely that a bad quality sample from a genuine user results in a high similarity score. Similarly, neither a good nor bad quality sample from an adversary is likely to generate a high similarity score with the template.

Using the NIST BSSR1 dataset, the authors conclude that min-max normalisation is the best performing common normalisation technique, albeit one that is sensitive to outliers in the data. In min-max normalisation, let X denote the set of raw matching scores from a given biometric expert and let the score $x \in X$. The normalised score is denoted x'. The method maps the raw scores to the interval [0, 1] and retains the original distribution of matching scores except for a scaling factor. Given the maximum and minimum values of the raw scores, the normalised score is calculated as:

$$x' = \frac{x - min(X)}{max(X) - min(X)} \tag{39}$$

For each of our experiments, repeated sub-sampling cross-validation is used to split the dataset into training and validation data. For each split, the model is fit to the training data to estimate the normalisation parameters and error-rates required by the fusion schemes. The predictive accuracy of the fusion algorithms are assessed using the validation data. The partitioning of scores are repeated 10 times and the results averaged over the splits. Figure 26 shows the receiver operating characteristic (ROC) curve for each of the individual experts, the sum rule and error-rate-based fusion. The results are summarised in Table 8.

Table 8 shows that the overall performance when using a combining scheme is better than any of the individual experts. The sum rule has an error rate of 1.11%, which is significantly lower than the best performing biometric expert. This increase in performance is expected and is in line with results published by Kittler et al.

Error-rate-based fusion has an error rate of 1.06% and thus outperforms the sum rule. This decrease in error rate stems from the fact that we weigh the scores given by each expert based on their FAR and FRR values as outlined in Equa-

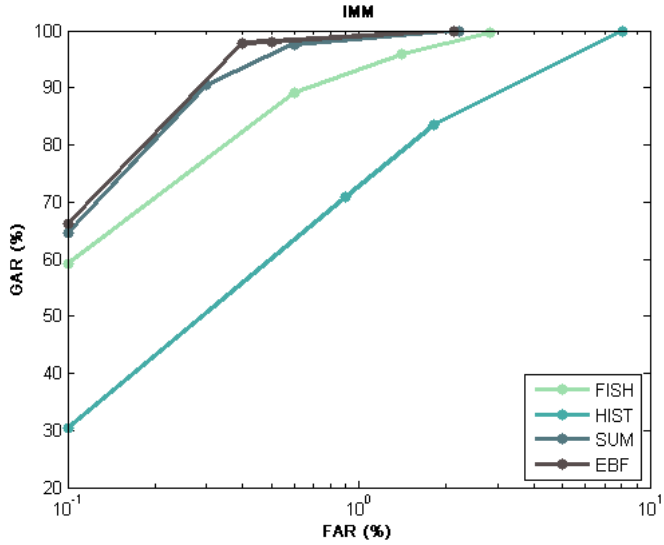


Figure 26.: Performance of Linear Discriminant Analysis and colour histograms on the IMM Face database

tion 19; conversely the sum rule uses fixed values for each w_i as given by Equation 38. As a result, our error-rate-based fusion strategy is more likely to resolve the conflicts between experts in favour of the best performing expert for the given biometric score.

Table 8.: Performance comparison of fusion schemes

Fusion Scheme	**TAR**	**FAR**	**ER**
Sum Rule	90.5%	0.3%	4.89%
	97.6%	0.6%	1.51%
	100%	2.22%	1.11%
Error-rate-based	97.8%	0.4%	1.31%
	98.1%	0.5%	1.20%
	100%	2.11%	1.06%

7.2.3 *State of the Art in Score Level Fusion*

We evaluate error-rate-based fusion on the NIST BSSR1 dataset and compare the results to the state of the art in score level fusion. From section 2.4 follows the division of score level fusion into three categories: (1) transformation-based score fusion in which match scores are first normalised (transformed) to a common domain and then combined, (2) classifier-based score fusion in which scores from multiple matchers are treated as a feature vector and a classifier is constructed to discriminate genuine and impostor scores, and (3) density-based score fusion, where scores are converted to likelihood ratios with an explicit estimation of the genuine and impostor score densities.

We compare the performance of error-rate-based fusion to a score level fusion scheme from each of the three categories, specifically: (a) a particle swarm optimisation (PSO) algorithm using z-norm normalisation [165], (b) a classifier-based score fusion scheme using support vector machines (SVM) [167], and (c) a likelihood ratio-based fusion scheme with Gaussian mixture model-based density estimation [39]. In short, PSO fusion finds the parameters for dynamic weighting and fusion of scores using particle swarm optimisation. This is a function of the accuracy and the degree of correlation between the biometric classi-

fiers. SVM fusion splits the training data into two classes with a hyperplane that maximises the margin between them. Scores from multiple matchers are treated as a feature vector and a classifier constructed to discriminate genuine and impostor scores. In likelihood ratio-based fusion the optimal combination of match scores are based on the likelihood ratio test. The distributions of genuine and impostor scores are modelled as finite Gaussian mixture models.

We assess the increase in true acceptance rate when using each of the fusion schemes, compared to the best performing individual expert. Further, to support the comparison and to provide a baseline for further use, we fuse the scores using the sum rule. As mentioned, the motivation for this approach arise as the performance of each fusion scheme is not only determined by their implementation but more so by their operating parameters. Fusion schemes may require complex pre-processing and normalisation techniques and detailed modeling of the score distributions. Also, to obtain the high precision rates in the published results, one must carefully select the operating parameters of each method. Therefore, there exists a trade-off between implementation complexity and precision, which is rarely reflected upon in literature.

The scores in the NIST face database are generated by two facial recognition systems, c and g, which are listed to output scores in the range $[0,1]$ and $[0,100]$. The score distributions are nonhomogenic, as seen in Figure 27. Note, that for system c there exists a small number of outliers in the data, caused when the system outputs a discrete score with the value –1. The other scores from the recognition system is in the interval $[0.35,0.95]$. These outliers have been omitted in the normalisation step.

The scores are processed and a threshold set for deciding whether to accept or to reject a match. A genuine match is obtained when two feature vectors corresponding to the same individual are compared, and an impostor match is obtained when feature vectors from two different individuals are compared.

The methods for dividing the NIST-BSSR1 dataset into training and test-sets vary greatly between implementations. Thus, in this evaluation we employ repeated sub-sampling cross-validation to split the dataset in an unbiased fashion. For each split, the model is fit to the training data to estimate the normalisation parameters and error-rates required by the fusion schemes. The partitioning is repeated 10 times and the results averaged over the splits. Figure 28 shows the receiver operating characteristic curves of the two face recognition systems,

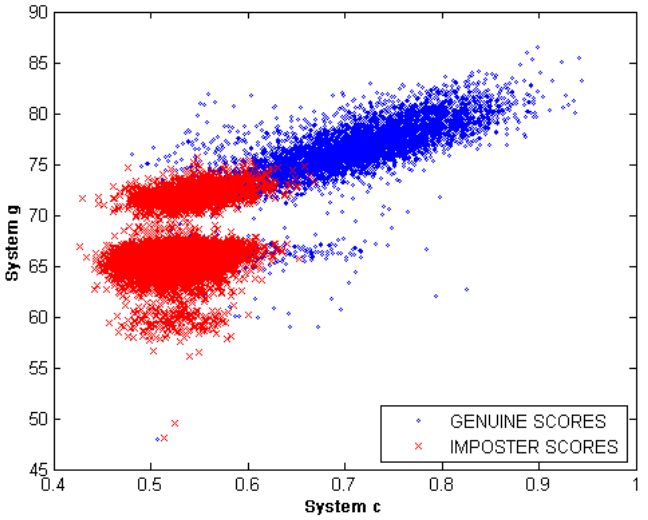


Figure 27.: Genuine and impostor scores for each of the two face matchers

error-rate-based fusion and the sum rule combining scheme. Error-rate-based fusion leads to significant improvement in the performance compared to the best single system and to the sum rule.

For each fusion scheme, we compare the reported genuine acceptance rates with the false acceptance rate set to 0.01%. The motivation for such a low FAR is that the false acceptances measure the portion of unauthorised users who are allowed in the system. In traditional applications of biometric authentication, where the system is used as the verification scheme for access control, even a single intruder is considered a serious threat to the overall security of the application. Note, that in contrast to these traditional applications of biometric authentication, persistent authentication tolerates a higher FAR, as the system continuously performs authentication of the principals. As such, wrongfully admitted principals may be detected and their authentication sessions revoked at a later stage.

Table 9 lists the performance comparison of the fusion schemes. The experimental results show that error-rate-based fusion outperforms support vector machine classification, particle swarm optimisation and the sum rule. Only likeli-

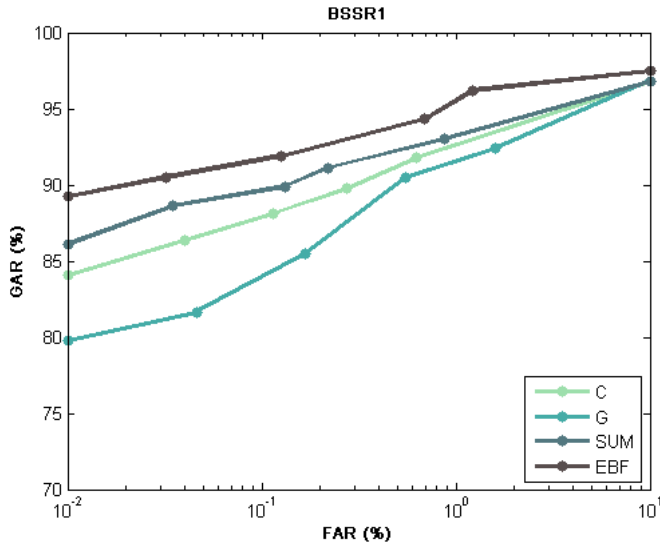


Figure 28.: Performance of error-rate-based fusion and sun fusion on the NIST BSSR1 dataset

Table 9.: Performance comparison of fusion schemes with FAR 0.01%

Fusion Scheme	Best Matcher	Fusion	Gain
Likelihood	71.20%	77.20%	6.00%
SVM	73.35%	77.50%	4.15%
Sum Rule	84.07%	86.08%	2.01%
Error-rate-based	84.07%	89.24%	5.17%
PSO	87.30%	89.50%	2.20%

hood ratio-based fusion performs better. Yet, the performance of error-rate-based fusion is close and in some applications error-rate-based fusion may be preferred over the Gaussian mixture model-based density estimation. Gaussians are not always appropriate for modelling biometric scores, as the score distributions may have more than one mode or require a great number of training samples to converge.

To put these results into perspective, consider the comparison of fusion techniques conducted by NIST [168]. With data from 187,000 subjects the authors evaluated eight biometric fusion techniques and their findings paint a similar picture. The authors concluded that density-based score fusion schemes consistently are the most accurate - but also the most complex to implement. This complexity is in the modelling of distributions, rather than in the fusion per se and the effect is especially true for density estimation at the tails.

For error-rate-based fusion this effect is evident as the performance of the fusion scheme diverges towards the tails. As is illustrated on Figure 28, the overall error rate of the fusion scheme at FAR 0.01% is 5.39%, whereas the best performance is achieved with FAR 1.22% resulting in an error rate of 2.51%. For persistent authentication, where the occurrence of false acceptances are mitigated by the continuous sampling of the biometric modalities, setting a higher FAR will lead to better performance in the fusion process, without causing undue security concerns.

The accuracy of error-rate-based fusion is significant and warrants a use in security sensitive biometric applications, where the performance of the biometric system is important. Error-rate-based fusion is especially suited for use with persistent authentication, as it integrates directly with the prior probabilities in the identity of the principals and benefits from the continuous authentications

provided by the framework. On the other hand, for less security dependent applications, a classifier-based approach may offer adequate accuracy. Further, as classifiers are simple fusion operators, they are significantly easier to implement and maintain.

7.3 EVALUATION OF THE TRACKER

The evaluation of the tracker first establishes the baseline performance of motion tracking and then assess the performance of restorative tracking with continuous authentications provided by the biometric experts.

The tracker is initialised on the first frame of the sequence that contains the principal and tracks the selected principal until a tracking failure occurs or the end of the sequence is reached. The generated trajectory is compared to the ground truth and the performance is measured as the number of frames the principal was successfully tracked. This measure is recorded as the number of frames where the overlap with the ground truth bounding box is larger than 50%.

7.3.1 *Evaluation of Motion Tracking*

We use the INRIA dataset to assess the performance of the tracker. A subset of the dataset is selected that contains people walking, browsing, meeting and interacting as a group. In this subset 32 unique principals are identified. The tracker is initialised on the first frame containing the principal and the measured position of each principal is noted and compared to the ground truth. The results are shown in Figure 29, which details the results for each of the 32 principals.

The figure shows the number of frames the principals are successfully tracked by the persistent authentication system (white) compared to the ground truth (red). The figure shows that the accuracy of the tracking algorithm, for most sequences, is very high.

For the first 18 principals there are few to no occlusions and no drop-outs, i.e., principals that are obscured completely from the view of the camera, and thus the system achieve near perfect tracking. In contrast, the remaining 14 principals are selected to test the performance in cases with heavy occlusions and regular drop-outs. In these sequences, the principals are often completely occluded or walk in

groups that make individual tracking challenging. This causes the tracker to rely on motion estimation to re-associate lost authentication sessions using the last know position of the principal and the associated trajectory and velocity. This is an error prone process that may cause tracking failures. Consequently, the accuracy of the system drops considerably in these clips.

The overall accuracy of the system on the CAVIAR INRIA dataset is 86.53%. While this might be sufficient for general positioning of the principals, it is too low for security sensitive applications. Thus in the following we evaluate the addition of remote biometrics for continuous authentication.

7.3.2 *Evaluation of Restorative Tracking*

We use the CAVIAR Lisbon dataset to evaluate the performance of persistent authentication with continuous verification. The dataset is selected as it provides a challenging setting that resembles the envisioned installation facilities. The biometric characteristics used in the evaluation are facial recognition and appearance analysis, both measured from a distance. Authentication is performed continuously by sampling the modalities recurrently, and the output of the biometric experts are fused using error-rate-based fusion.

Like the INRIA dataset, 32 unique principals are identified and tracked to test the performance of persistent authentication with fused biometrics. The tracker and the remote biometrics are initialised on the first frame in the sequence that contains the principal and has available biometric modalities. The performance of the system is measured by recording the number of frames each principal has been successfully tracked by the persistent authentication system compared to the ground truth. To evaluate the effect of the remote biometrics we assess the performance both with and without continuous authentications and compare the results.

For continuous authentications, we sample the biometric features as they become available. The setting in a corridor, where the principals are walking in both directions, means the principals are not always facing the camera. As a result, the experts are only able to extract features from a subset of the total frames. In addition the results may be less accurate due to occlusions and changes in orientation.

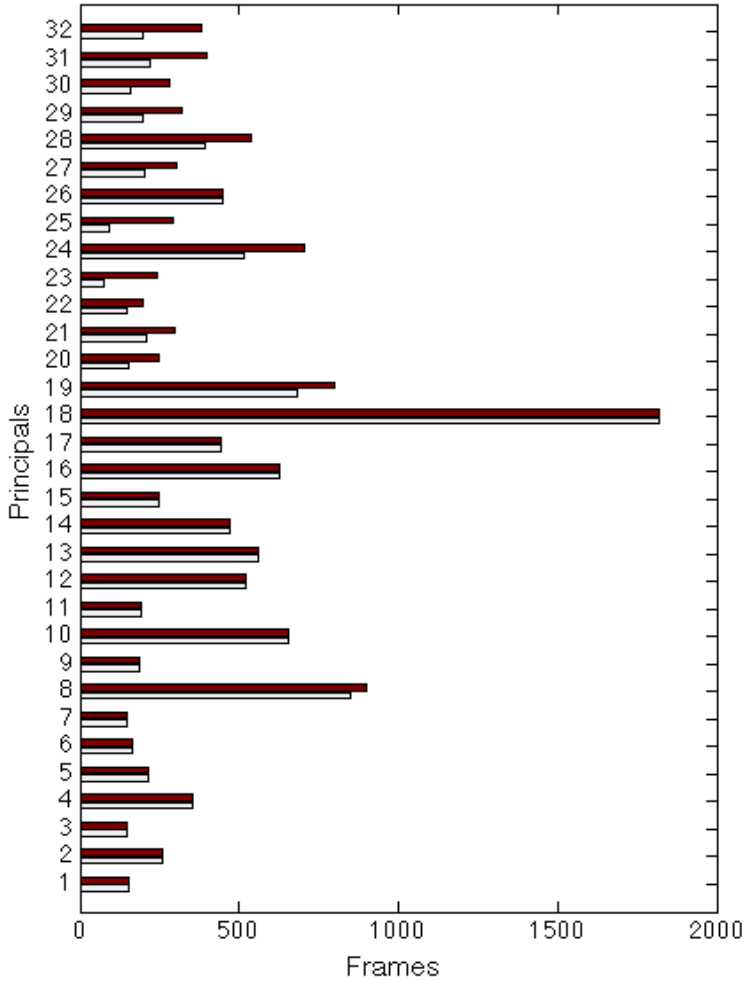


Figure 29.: Frames the principals are tracked by the persistent authentication system (white) and the corresponding ground truth (red)

Figure 30.: The varying poses and sizes of the captured faces from the CAVIAR Dataset.

The 32 principals are tracked through varying poses, distances to the camera and in changing illuminations. An example of this variance is shown in Figure 30 for three principals. The half-resolution PAL standard, which the CAVIAR videos are recorded in, are considered quite low by todays standards and thus the resolution of the captured facial images are very low. The examples shown in Figure 30 are only 50 x 50 pixels, which we conjecture will have an impact on the performance of the experts.

The dataset is constructed from the initialisation sample of each of the tracked principals, as well as faces from 18 other occupants which provide additional imposter scores. The normalisation parameters and error rates are calculated and for each subsequently captured biometric sample this process is repeated using leave-one-out cross-validation. New samples are then added to the training set to increase the robustness of the system. In contrast, for a production system, the training set would be constructed beforehand, using high quality samples captured from each principal during the enrolment process.

Each step is monitored by a human expert who records the performance of the system and of each of the biometric experts. Figure 31 shows the receiver operating characteristic curves of the face recognition system, the colour his-

tograms, error-rate-based fusion and the sum rule combining scheme. The FAR value yielding the best performance for each system is summarised in Table 10.

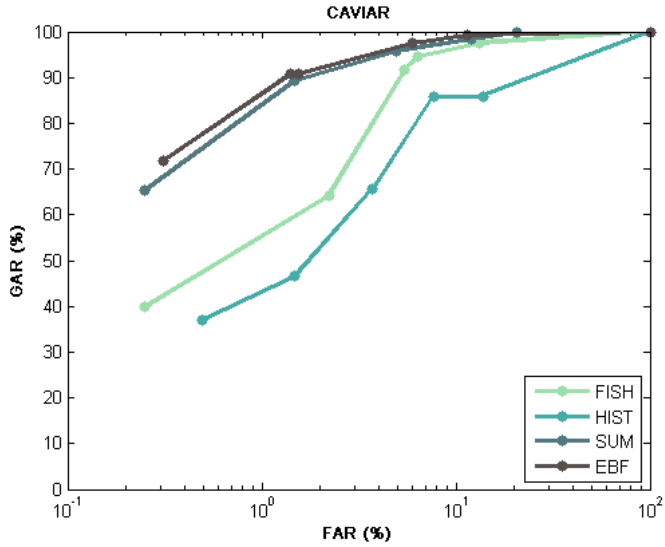


Figure 31.: Performance of error-rate-based fusion and sun fusion on the CAVIAR Lisbon dataset

Table 10.: Performance of Biometric Experts and Fusion Schemes

System	TAR	FAR	ER
Facial Expert	94.58%	6.40 %	5.91%
Apperance Expert	85.96%	7.64%	10.84%
Sum Rule	95.81%	4.93%	4.56%
Error-rate-based	97.50%	5.98%	4.24%

Error-rate-based fusion performs significantly better than the best single system and also outperforms the sum rule. Error-rate-based fusion has an error rate of 4.24%, compared to 5.91% and 10.84% for the two biometric experts. Likewise, the sum rule performs really well on this dataset with an error rate of 4.56%. We conjecture that the relative high error rate of the individual experts, especially the facial expert, on this dataset is caused by the very low resolution of

the training images and the greatly varied poses of the principals. However, this shows that even in adverse conditions the LDA method yields usable results. Additionally, the simple approach of the appearance expert is still able to provide meaningful information for the fusion algorithms.

With the performance of the biometrics determined, each of the 32 principals are tracked following the framework outlined earlier. The number of frames each principal is successfully tracked is measured and compared to the ground truth. This process is repeated twice, once for motion tracking and once for restorative tracking.

The results are shown in Figure 32, which charts the results for each of the 32 principals. The figure shows the number of frames the principals are successfully tracked by the motion tracker (white) with remote biometrics (grey) and the ground truth (red). For some principals there are few or no occlusions and no drop-outs, and in these situations both systems achieve near perfect tracking of the principals. The accuracy of the tracking drops when occlusions and drop-outs occur, for instance when principals enter a shop or when multiple principals crowd the scene. The persistent authentication prototype may completely lose track of a principal and in this case the remote biometrics are used to re-associate the session with the correct principal. As a result, the system using remote biometrics greatly outperforms motion tracking for a number of the tracks. This effect is most pronounced for tracks 7, 13 and 21.

The overall performance of the system with only motion tracking is 73.60%, whereas using continuous authentication with remote biometrics results in an overall performance of 91.78%. Continuous authentication makes it possible to re-identify principals who are otherwise lost, due to occlusions in crowded scenes or drop-outs caused by camera positioning. The evaluation shows that the persistent authentication system is able to achieve good tracking results, even with low quality video data, in a challenging usage scenario.

7.4 SUMMARY

The persistent authentication prototype is evaluated on four publicly available datasets to ensure the reproducibility of the results. The datasets considered are, the IMM Face Database, the NIST BSSR1 dataset, the CAVIAR dataset from

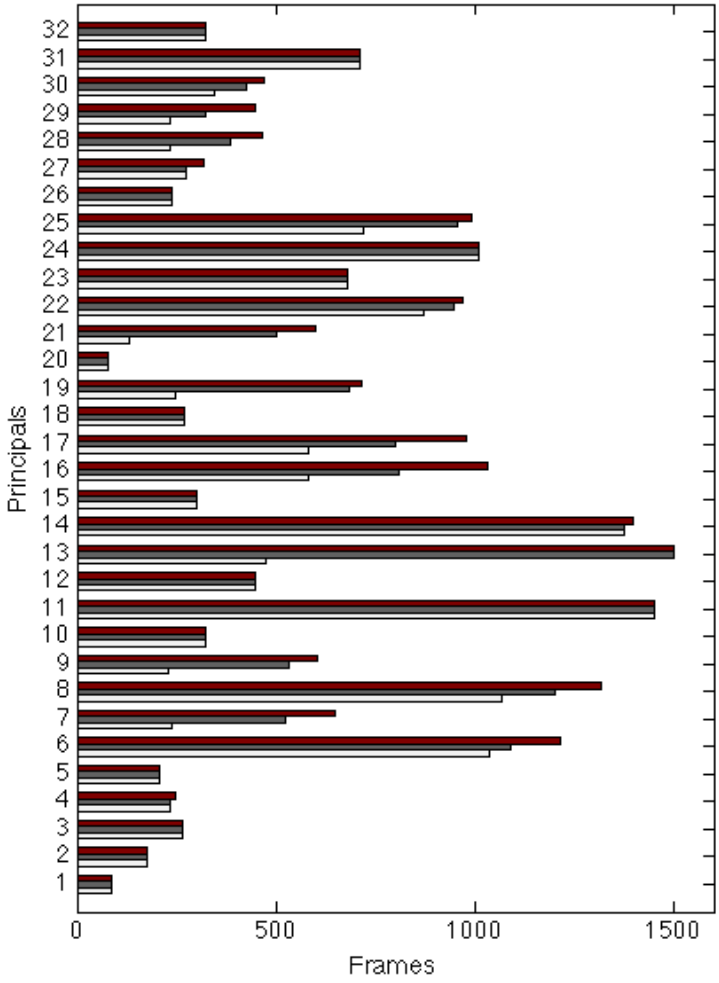


Figure 32.: Frames the principals are tracked by the persistent authentication system (white) with remote biometrics (grey) and the corresponding ground truth (red)

INRIA Labs at Grenoble in France, and the CAVIAR dataset from Lisbon in Portugal.

Two biometric modalities are considered, namely, facial recognition and appearance analysis. Both characteristics are measured from a distance with continuous authentication performed by sampling the modalities recurrently. The biometrics are used to evaluate the performance of error-rate-based fusion in comparison to the sum rule to provide a baseline for further evaluations.

The experimental results show that error-rate-based fusion outperforms support vector machine classification, particle swarm optimisation and the sum rule. Further, the performance of error-rate-based fusion approaches that of likelihood ratio-based fusion. In addition, for some applications error-rate-based fusion may be preferred over the Gaussian mixture model-based density estimation used in likelihood fusion.

Error-rate-based fusion used in combination with persistent authentication increases the tracking performance greatly. The persistent authentication prototype with continuous authentication using fused remote biometrics achieves 91.78% accuracy on the Lisbon CAVIAR dataset. Given the low quality video data and the limited training set, this shows the robustness of the persistent authentication method.

8

CASE STUDY

8.1 THE OBJECT

In the scope of the RIBS project a test facility has been provided for the partners to evaluate their protection measures, in order to ensure the convergence of the efforts and integrations. The test facility represents a generic commercial building located in Europe and is known as *the object*. Deploying persistent authentication as a protection measure in the object were a truly multi-disciplinary effort, contingent on the expertise provided by all the parters in the RIBS project, to whom credit is due for their invaluable help during the planning, installation and evaluation.

The object is a three story building with multiple service functions including both public interfaces, private interfaces and back offices. This evaluation primarily focus on the public interface, which is located on the ground floor. The public interface concerns the main function of the object, namely, the operation of a retail bank branch. The bank branch provides a number of services, including a customer front desk, automated teller machines (ATMs), meeting rooms for private conversations and waiting areas for the customers. There is considerable interaction between the public and private interface of the bank and the areas open to the public intersect the private areas and the back office. This creates an interesting setting for persistent authentication, as the layout of the public and private interfaces requires a disproportionate number of authentication points to secure. This reduces the usability of the facility and require the bank's employees to continuously authenticate to access the private areas of the bank.

The formal security practices in the object are fairly straightforward–there is a single entry point to the public section, beyond which smart-cards or keys are required. Employees pass through the same security system as visitors to the branch (double-doored turnstiles equipped with cameras, allowing only a single principal to enter at a time). Entry to the private interface is regulated with smart-cards. There is a single authentication zone at the entrance to each floor

beyond the ground floor, which gives access to the entire private section on that floor. Predominantly, the access control systems used in the object are in place to prevent 'casual' faulty entry and not made to withstand more deliberate or equipped malevolent entry. We would like to remind readers that the smart-card based authentication system, used for the initial authentication of the principals, is external to the persistent authentication system and as an interchangeable part of the system, we do not consider it directly.

Surveillance routines in the object rely heavily on passive/collective surveillance. Security guards are present at the entrance, beyond which most surveillance is done by personnel with other tasks (cashiers, consultants, advisors). The structure for this passive surveillance is eased by the generally sociofugal configuration of the object, i.e., that the functions in the object are located in the periphery, thereby concentrating visitors to central locations that are easier to monitor. Video surveillance are only in select places in the object, and mainly used to record events so as to establish proof and information for police investigations post-event.

The persistent authentication system were installed in the object over a period of two weeks. To substantiate the collected data, additional manual observations were carried out that recorded the location and activity of people in the object. These observations were followed up by interviews with the principals to discover additional contextual information, such as the frequency and reasons for their visits to the bank. These interviews were used in conjunction with the persistent authentication data to help detect behavioural patterns that encompass more contextual information than sensors alone can provide.

The location-based service considered in this evaluation is a context-aware access control system, similar to the one described in section 4.4. The initial authentications were provided by the smart-card system already in place in the object. The existing video surveillance were too sparsely distributed to provide contextual awareness to the persistent authentication system, thus additional CCTV cameras were installed in the object. To reduce the intrusiveness of the installation the number of cameras were limited, conversely, the objective of the installation was to maximise the coverage of each camera.

The CCTV cameras used in the installation are Level One network cameras[1] with 360-degree fisheye lenses. The 360-degree field of vision allows a smaller number of cameras to cover a larger area. On the other hand, the lens introduces distortion, especially in the periphery of the image. The CCTV cameras were mounted as recommended in the specification–in the ceiling, looking directly down. This provides a viewing angle that helps reduce the occurrence of occlusions, but makes it very difficult to capture faces of the principals. Consequently, the persistent authentication system relies on the existing video surveillance cameras and the smart-cards for continuous authentication.

The installation point of each camera and their respective fields of view are shown for the ground floor of the object in Figure 33- The radial distortion of the lenses were corrected using the methods outlined in subsection 6.6.1, and the correlation between overlapping areas, covered by multiple cameras, adjusted manually. Figure 34 shows the resulting area covered by the cameras. The purple line indicates the field of view of the cameras after distortion correction. The green line shows the field of view after taking into account the topology of the building and interposing objects. The total area covered by the CCTV cameras is 350 square meters.

To get a visual understanding of the distribution of people in the object, consider Figure 35. The figure shows the location of people over a two hour period, taken as five minute snapshots. This gives an indication of the areas with high traffic and the areas where people queue up.

1 The CCTV cameras have a 1/3" Progressive CMOS 2-Megapixel Sensor, recording at VGA resolution (640x480 pixel) using the H.264 compression format.

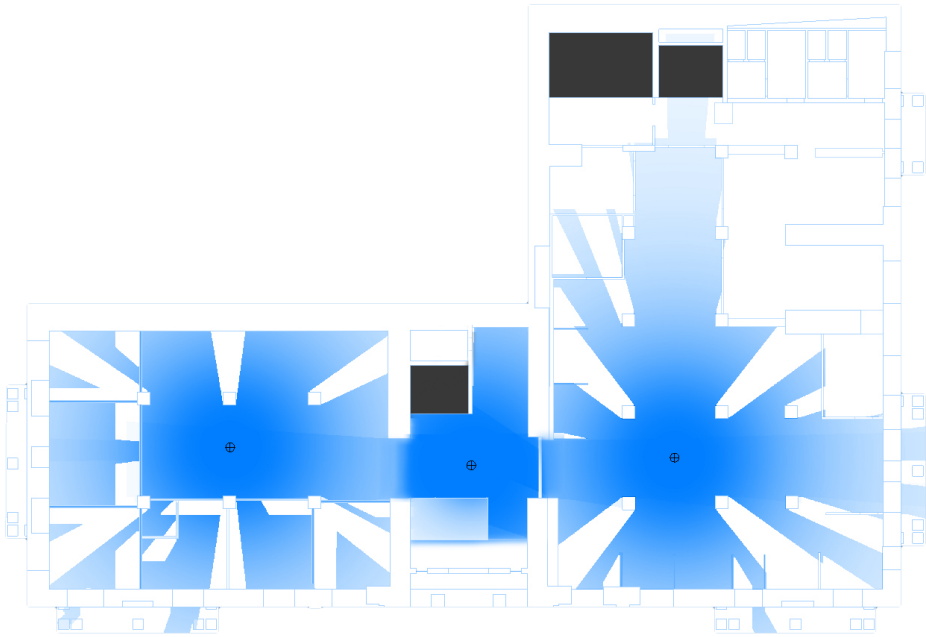


Figure 33.: Field of view of the installed CCTV cameras

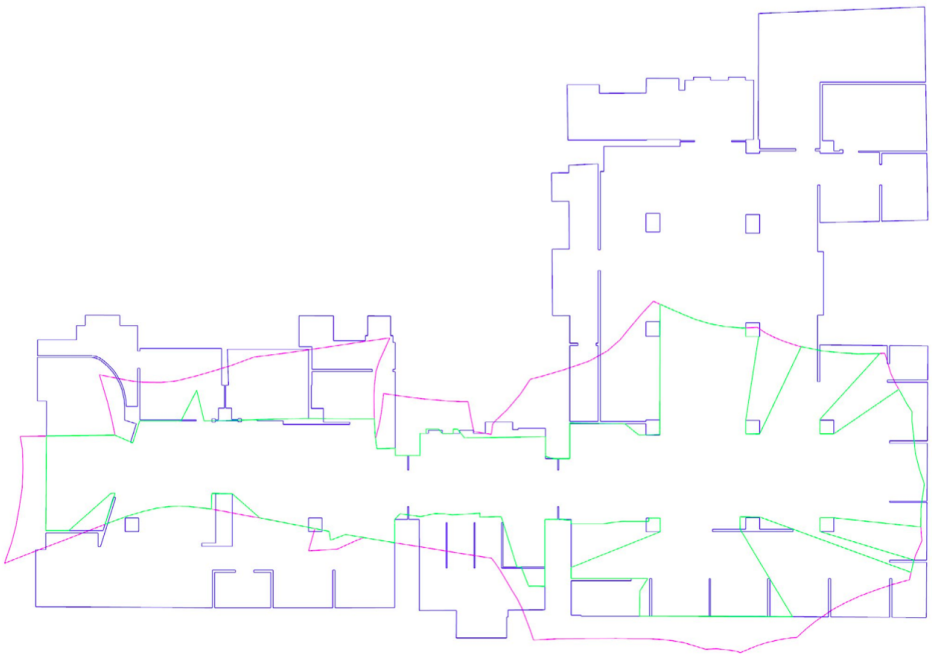


Figure 34.: Area covered by CCTV cameras after radial distortion correction (purple) and taking into account the topology of the building (green)

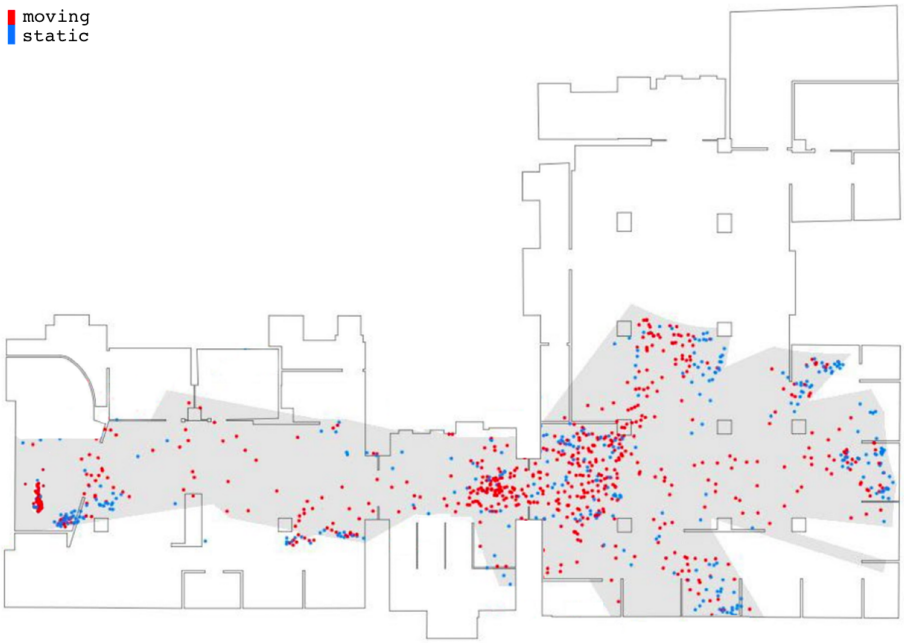


Figure 35.: Distribution of people moving (red) and waiting (blue) taken over two hours, as five minute snapshots.

8.2 SCENARIOS

The method adopted in the RIBS project to assess the proposed protection measures belongs to a group of requirements engineering methods known as scenario-based validation [169]. Scenario-based validation is an analysis that spans a given period of time, with the objective of testing a range of conditions that the protection measures may face. The relevant conditions are captured in a set of scenario models that represent the successive states of the system over the period.

In order to develop the scenarios our analysis starts with the *central event* [170]. In every attack there is a number of elements that are decided by the attacker when planning and executing the attack. The central event is identified as the defining element, that has the potential for causing severe consequences to the stakeholders, and are considered as the basis of the perpetrator's operational objectives.

The central events that are consider in the RIBS project are as follows:

- Activation of a biological weapon near or within the object
- Activation of a chemical weapon near or within the object
- Activation of an explosive weapon near or within the object

The constraints of the environment may introduce restrictions on which options are available to an attacker. As an example, the offender may choose to hide the weapon in a car. The presence of the weapon in a car creates a number of requirements on the scenario. For example, that the car must accommodate the weapon and needs to be moved to reach its intended location.

When developing new security technologies, it is useful to know how offenders might proceed to achieve their objectives. This information can be used to identify relevant factors of potential attacks and devise strategies and control principles to disrupt them. To this aim, attack models can be constructed in the form of crime scripts [171].

Thus, in order to create the scenarios a diverse range of operational elements, reflecting a variety of offender decisions, are selected and the scenarios interpolated from these. For the majority of the scenarios, the most direct conditional

path that meets the offender's objectives are selected, as it is assumed that the offender follows rational decision making theory and the least effort principle.

8.2.1 *Crime Scripts*

In the crime science literature, crime scripts are generally specified in natural language [171] [172]. The purpose of the following scripts is to test the adaptability of the persistent authentication system, i.e., to test if the system is able to detect the offender in the chosen scenario.

Scripts can never be at the same time completely generic and consistently applicable, which is why the RIBS project proposes a range of scripts that have a wide range of uses. A total of 12 crime scripts are defined in the RIBS project, most which deal with the central events and post-activation of a weapon in the object. In the persistent authentication system we are instead interested in the events leading up to the activation and the factors contributing to the delivery and propagation of the weapon by insiders and intruders. Thus, in the following we present a subset of the crime scripts, containing the two most applicable to the persistent authentication system.

STORAGE
Weapon E: 5 kg IED (improved explosive device)
"A cleaner will attempt to introduce a weapon in the bank, bringing in a small amount of agent at a time over several days. She intends to place the device in a board room on the office floor and trigger it remotely using a mobile phone, at a time where specific or many employees are in the bank." .

PACKAGE
Weapon E+B: IED + Bio agent
"An external supplier will attempt to introduce a small package, containing a device with a small explosive charge and a threat agent on the office floor. After leaving the bank, the individual intends to trigger the weapon using a mobile phone, and contaminate the employees."

The first script evaluates that the persistent authentication system can detect insiders, i.e., principals that are authorised to access parts of the object and can exploit these access rights to introduce an explosive device into a sensitive area. The second script evaluates that external personnel that are granted access to the object are detected if they attempt to introduced a combined explosive and biological weapon into a restricted area.

8.2.1.1 *Storage*

For the first script a set of logical security policies are created in the persistent authentication system, which define the areas authorised personnel are allowed access to. These policies include important appurtenant contextual information, such as the time of day and the length of access. The policies are formulated as detailed in section 4.4 and enables the persistent authentication system to detect insiders if they violate these security policies.

The virtual walls used in the security policies are created as outlined in subsection 4.4.2, with three separate entities defined. The first virtual wall covers general access to the office floor and monitors that only authorised personel enters the back office. Entry to the back office is regulated with a smart-card based access control system that were already in place in the object. The second wall covers access to the board room, which likewise uses smart-cards for security. Finally, the third virtual wall covers entrance into a manager's private office. Access to this office is not specifically restricted beyond the general access control to the back office, but entry is only intended during office hours, when the managers is present. We specify the walls as follows, with the response represented by a boolean, *accept/reject* based on the evaluation of the principal:

$$\begin{aligned} w_1 &= \langle \textit{back office, employees, response} \rangle \\ w_2 &= \langle \textit{board room, employees, response} \rangle \\ w_3 &= \langle \textit{private office, manager / \{employees\}, response} \rangle \end{aligned} \tag{40}$$

To assess the robustness of the security policies we monitor and log access to the back office, the board room and the private office. We note the entries detected by the persistent authentication prototype and compare it to a manually annotated ground truth. The performance is measured as the fraction of correctly measured events detected by the prototype out of all events. Table 11 summarises

the result of the evaluation over a 8 hour data collection period, for each of the three virtual walls.

Table 11.: Evaluation of virtual walls in the object

Virtual Wall	Entries	Genuine	Measured	FRR	Performance
Office floor	231	207	227	1.73%	98.27%
Board room	52	52	45	13.46%	86.54%
Private office	28	28	23	17.86%	82.14%

Out of the 231 entries to the office floor the majority, 207, are legitimate entries in which the personel correctly presented their smart-cards before entering the back office. In the remaining entries, employees were tailgating their colleagues through the access control system in order to avoid the hassel of authenticating themselves. The persistent authentication prototype successfully detected 227 of these entries including all the illegitimate entries, giving an overall performance of 98.27% with no false acceptances. The high detection rate is in part due to the proximity of the smart-card authentication system to the virtual wall and in part due to the sparse population of the access area. This provides optimal operating conditions for the persistent authentication prototype, which is reflected in the detection rates.

In the 52 entries to the board room the persistent authentication prototype correctly detected 45 entires, resulting in an overall performance of 86.54%. Only genuine access occurred, thus the false rejection rate was 13.46%. The lower accuracy of this result compared to entry to the office floor, is due to the way people access meeting rooms. People will generally queue up and enter the board room in groups, creating a very challenging situation for the prototype where identifying each individual in the group is difficult. This causes the prototype to lose track of the authentication sessions associated with each individual.

An extension to the persistent authentication prototype is proposed to deal with groups of authenticated people by merging authentication sessions using the least common denominator for each group. For instance, a group of authenticated principals queuing up together are all authorised to access restricted areas based on the individual in the group with the lowest privileges. Thus, even in cases where individual tracking is impossible, this allows the system to authorise principals where the service provision threshold, Δ_j, of the authorisation zone,

S_j, is lower than the least common denominator of the group. This approach increases the usability of the system and is especially suited for areas, such as meeting rooms, that do not require non-repudiation.

For entry into the private office the persistent authentication prototype correctly identified 23 out of the 28 entires. No illegitimate access was attempted and the overall accuracy were 82.14%, which is considered quite low for the prototype. One of the explanations are that access to the private office requires the prototype to correctly evaluate the credentials of both the manager and the accompanying employees. Further, the employees may authenticate at the entry to the office floor and then wait before going to the managers office. Hence, the persistent authentication prototype must track these principals for an extended period of time, before their authentication sessions are needed. As the smart environment is not specifically setup to facilitate acquisition of remote biometrics the number of positive biometric signatures are quite sparse. This has an adverse affect on the confidence of the tracker, and ultimately results in poorer performance. In environments where installation of more sensors are possbile, the performance of the persistent authentication system will increase.

The demonstration shows that the persistent authentication prototype is able to detect access and to determine the credentials of the involved principals. However, the static nature of the virtual walls and the security policies means that attackers are only detected if the policies are violated directly. For instance, if the cleaner is allowed access to the board room at a specific time and uses this time to introduce the weapon, the persistent authentication system is unable to detect the attack. It is therefore necessary to introduce dynamic constraints in the formulation of the security policies.

Such dynamic constraints uses the movement patterns of the occupants to create behavioural profiles based on data analysis of historical data. These behavioural profiles enables the persistent authentication system to detect if the movement of an attacker deviates from the established patterns, either the patterns created by all occupants or specifically the patterns created by the cleaning personnel. Further, it is possible to detect deviations in the patterns generated by the attacker over time, i.e., detecting if the attacker shows any unexpected changes in behaviour. In the script, these dynamic constraints can facilitate the detection of the attacker if the introduction of the IED to the board room causes sufficient deviations from the normal patterns.

We evaluate the detection using dynamic security policies by considering different normal patterns of the object. We learn these normal patterns by aggregating the output of the persistent authentication tracker. Figure 36 shows an example of the movement patterns generated by the occupants of the object. Patterns that deviate from this established norm can be detected by the persistent authentication system. The deviations can be measured using metrics such as velocity, access time, density at location, and more.

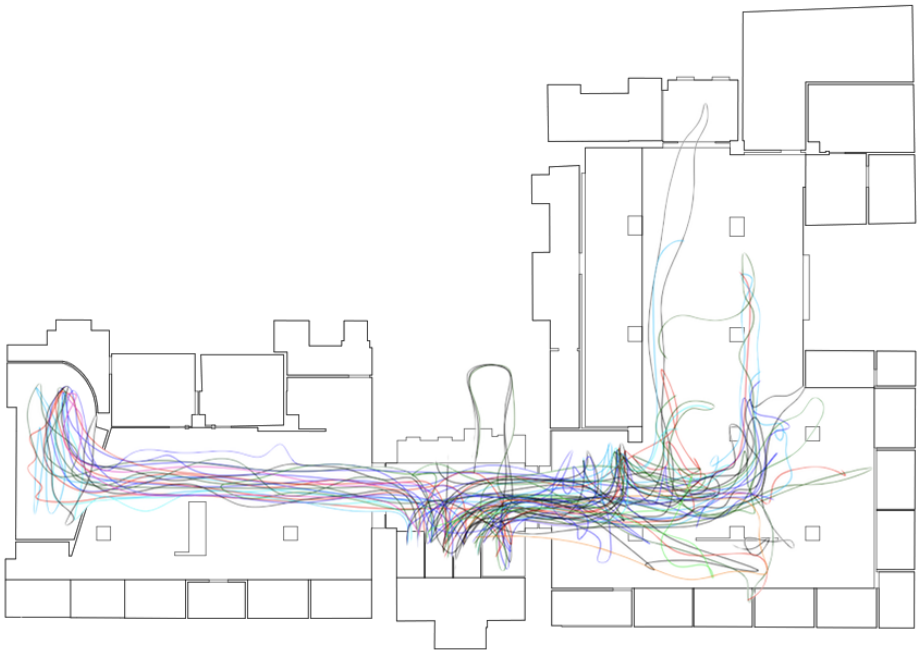


Figure 36.: Example of the tracks generated in the object

The performance and viability of this approach are determined by the detection rate in contrast to the number of false positives. This primarily depends on the rigidity of the detected movement patterns. Thus, for an organisation with very specific working conditions and a methodical approach, dynamic constraints may offer a good detection rate with few false positives. Conversely, in an organisation with diffuse working conditions the approach may not be suited to detect insiders, and instead other measures, such as stricter surveillance or tighter

access control, are needed. In either case, detecting insiders is one of the most challenging tasks to solve for surveillance applications and ultimately detection depends on the actions of the attacker and the specific scenario.

Figure 37 shows a few examples of the deviation from normal patters that were detected in the object. These include (a) groups of people loitering in access areas, (b) persons running or making quick dashes to catch an elevator, (c) impromptu meetings between the personnel, and (d) cleaning personnel moving furniture at night. These examples help illustrate that deviations can be detected, but are primarily useful as a warning system, that requests human attention.

Figure 37.: Deviations from normal patters

8.2.1.2 *Package*

In the second script the attacker is an external supplier, thus no prior information is known about the attacker's movement patterns. Instead the persistent authentication system relies on tracking the movement of the supplier to detect any precursors to an attack. The supplier is authenticated at the front desk by stating the intent and recipient of the package. The supplier is allowed entry to the private interface of the bank on the condition that the delivery goes directly to the intended recipient through a specific route. Within the persistent authentication system this is considered a conditional authentication session, i.e., an authentication session that may be revoked based on the actions or contexts of the principal. The persistent authentication system monitors the authentication session and if the conditions are no longer met an alarm can be raised.

Conditional authentication sessions are useful for external personnel or guests visiting restricted areas, which both pose critical security concerns. This is illustrated by an incident that occurred in the national parliament of Denmark in

2003. Two members of a grassroots movement entered the parliament under the guise of visiting a minister. Instead they made their way unchallenged to a foreign policy meeting and assaulted the prime minister and the secretary of state with red paint in frustration over the Danish participation in the Iraq war. With conditional authentication sessions and the tracking capabilities offered by the persistent authentication system this behaviour is detectable and actionable.

8.2.2 *Behavioural Patterns*

In persistent authentication we detect insiders, intruders, and hostile reconnaissance by learning the contextual and behavioural patterns generated by the movement and interaction of principals in the environment. We use the state of each principal and the flow techniques discussed in section 6.1 and section 6.3 to create aggregated movement patterns of both authenticated and unauthenticated principals. While we conjecture that attacks are distinguishable from normal patterns in a suitable feature space, we observed no malicious behaviour during the two week data collection period in the object. Thus, we will in this section instead focus on some of the more benign behavioural patterns we detected and how these integrate with the operation of the bank branch. In the following we present our findings and experiences with the system and the tools we have developed to identify the dynamically changing behavioural patters.

Surveillance systems generate vast quantities of data. This data must first be processed and then analysed before the prevalent patterns can be identified. Motivated by this we have developed a dedicated data exploration tool in collaboration with AEDAS that is capable of processing and structuring very large quantities of data. We integrate the processed data with state of the art data visualisation tools to present the information in a meaningful way.

The data processing tool is written in Java and presents the user with a top-down overview of the facility as seen in Figure 38. The tools loads an *.dsf* formatted file that describes the structure and layout of the facility. The tool then automatically creates partitions, such that each room in the facility is identified in the model. The tool also takes the camera placements and viewing angles described in section 8.1 as an input. Data is loaded in YAML-files, containing state

information from the persistent authentication prototype in each frame of the sequence.

The tool is used for data exploration and presents the user with a histogram of the historic data (second image in Figure 38). At a glance this gives an indication of the level of occupancy in the facility. Each room in the model can be selected and the histogram updates to reflect only the chosen partition. Dragging the slider processes the data in chronological order and allows the user to quickly examine the data. In addition, the tool provides the option to calculate the density of occupants in the facility. This is illustrated in the third image in Figure 38 where the heatmap colours indicate either a high or low density. This feature is useful to detect choke points and areas of interest in the data. The tool is meant to quickly allow the user to form hypothesises about certain patterns in the data. Different hypothesises can then be investigated and the data foundation for each recorded in the tool. The selected data can then be exported for later analysis and visualisation.

In the following we present our findings from the case study in the object. In the identified patterns there is a clear correlation between the movement of principals and the spatial integration of services. Additionally, for the majority of the patterns there is a strong component of time dependent factors and relation to specific types of visits, which illustrates that the behavioural patterns contain strong social structures and vice versa. Specifically we have found:

- Visits to the bank branch consist of a mix of explorative and purposeful visits, i.e., some visits serve to gather information on contextual factors such as waiting time, whereas other visits have a clear functional purpose such as interaction with the branch functions.

- The movement flows are strongly connected with areas that allow for a good overview of the functions and services provided in the object. These areas are of further interest as they are prime spots to perform hostile reconnaissance.

- The sociofugal configuration of the object creates a strong distinction between accessing or waiting for access to the branch functions and interfaces. Thus, dependent on the motive of an attacker, the detected patterns provide different potential that are located very differently in the object.

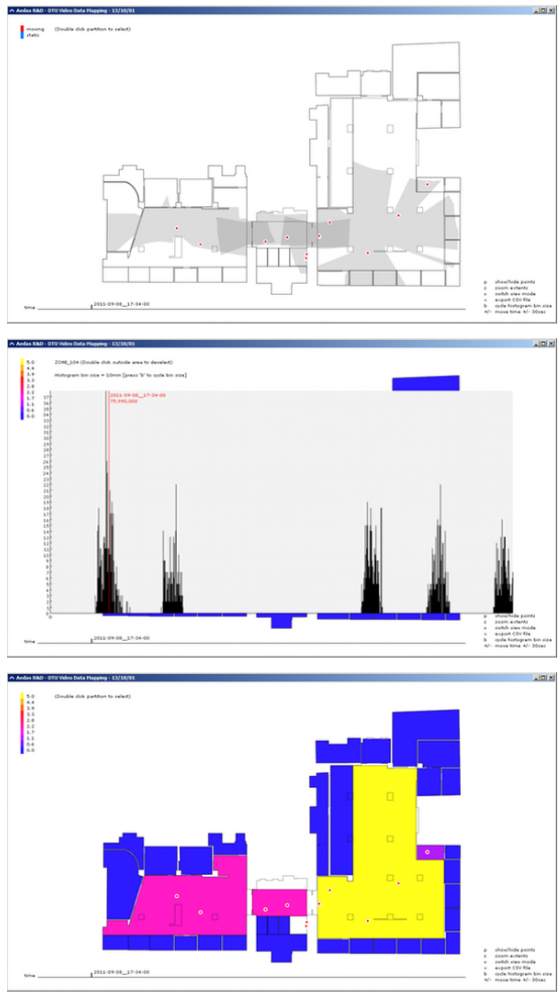


Figure 38.: Data processing tool

- The patterns generated by the employees of the branch have little effect on the overall movement flow. However, it has high impact on occupancy, specifically the degrees of staff presence in different parts of the object.
- The automated interfaces have a stricter relation between occupancy and movement patterns than the personal interfaces. This is likely due to, on one hand the service provision time, and on the other the social character of the personal interface.

We found that many of the visits to the bank serve to survey the current state of the object. That is, visits with the purpose to see how long the line is to a certain function, or to see if a personal contact is available. This is pointed out as a deviation from normal behaviour, as it is a common perception that people only enter organisations, especially banks, to make use of some of their services. Contrary to this notion, nearly half of the visits to the object exit without interfacing with personnel or service provision points. An example visualising this behaviour is shown on Figure 39. The image series show how the movement flow evolves with the availability of service provision points. The first frame shows the general flow for a specific section of the object. In the second frame an additional personal service point opens, which, in the third frame, creates an increase in the flow to the area, as people begin to make use of the newly opened service. In the fourth frame we see a circular pattern in the flow, created by principals who, after surveying the waiting time to the service point, decide to postpone their visit and leave the bank.

Another interesting behaviour we found is that tailgating through the access control points are common amongst all employees. However, it should be noted most employees know one another and would recognise if a stranger attempted tailgating. A more serious problem is that certain deliveries leave the doors to the private interface open when delivering goods. As there is only a single access control point, this compromises the security of the entire floor. Persistent authentication can be used in both cases to detect if unauthorised principals entered the restricted area, either through tailgating, forced entry or by slipping in during a delivery.

Finally, the interaction between the spatial configuration of the object and the movement patterns of the occupants are analysed. The analysis is visualised by processing the dataset and recording each unique position, which corresponds

to the occupancy density for that particular point. The results are shown in Figure 40. The figure clearly illustrates that some areas have high densities and that other areas are mostly used as access routes. The areas with high densities are characterised by the presence of important service points, such as cashiers, ATMs, and elevators.

The visualisation proves relevant not only for spatial analysis but equally for understanding where the object's generic interface reveals pressures at certain times of the day and how the spatial arrangements become either overloaded or underused. This can be of vital importance when it comes to evacuation or emergency situations, or in the identification of choke points and high density areas, that are prime targets for attacks and reconnaissance.

8.3 SUMMARY

A test facility has been provided for the partners in the RIBS project in order to ensure the convergence of the efforts deployed as part of the project. The test facility represents a generic commercial building located in Europe and is known as *the object*. The object is a three story building with multiple service functions including both public interfaces, private interfaces and back offices. The public interface concerns the main function of the object, namely, the operation of a retail bank branch.

In the evaluation the movement patterns and behaviours of the occupants in the object are explored. We found that depending on the motive of an attacker the detected patterns provide different potential that are located very differently in the facility. The scenario-based validation shows how persistent authentication can help detect insiders by leveraging static security policies and the dynamic behavioural patterns. These patterns provide important prior information that enables the persistent authentication system to detect deviations in the behaviour of occupants in the facility. In addition the scenario-based validation shows how conditional authentication sessions, based on tracking data, can be used to detect malicious external personnel.

A data processing tool is developed to quickly allow the user to form hypothesises about certain patterns in the data. Different hypothesises can be investigated and the data foundation extracted. The data integrates with our visualisations,

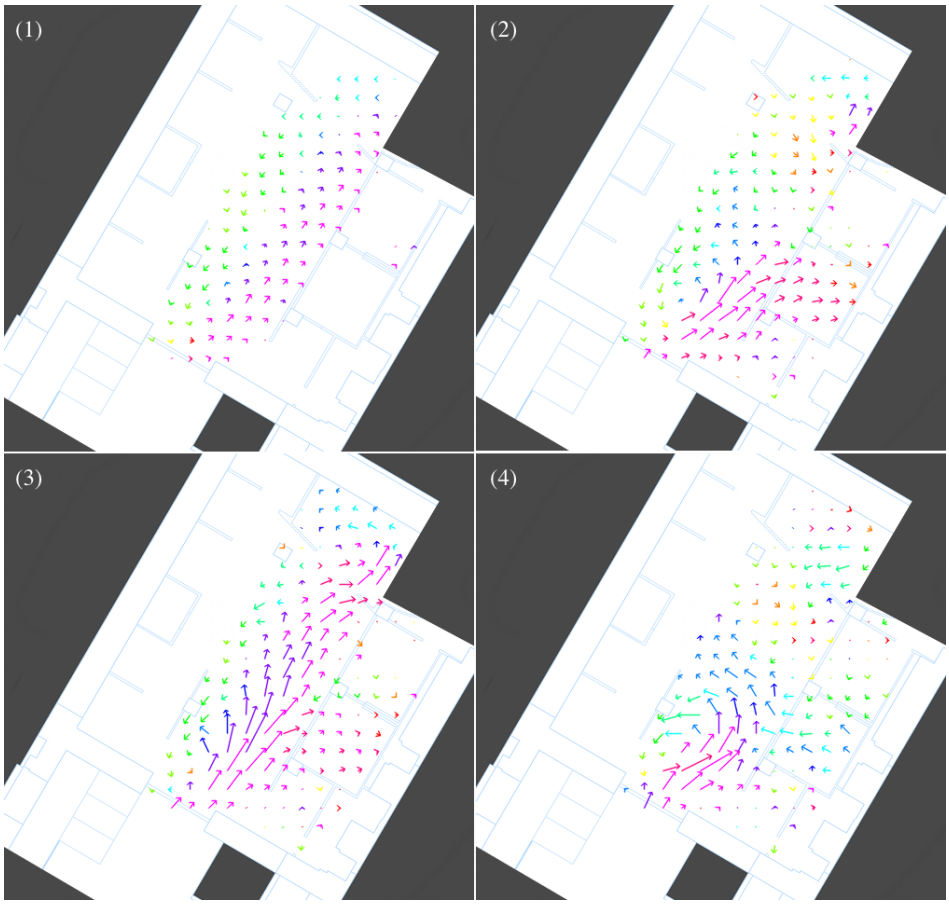


Figure 39.: Behavioural patterns in the movement flow

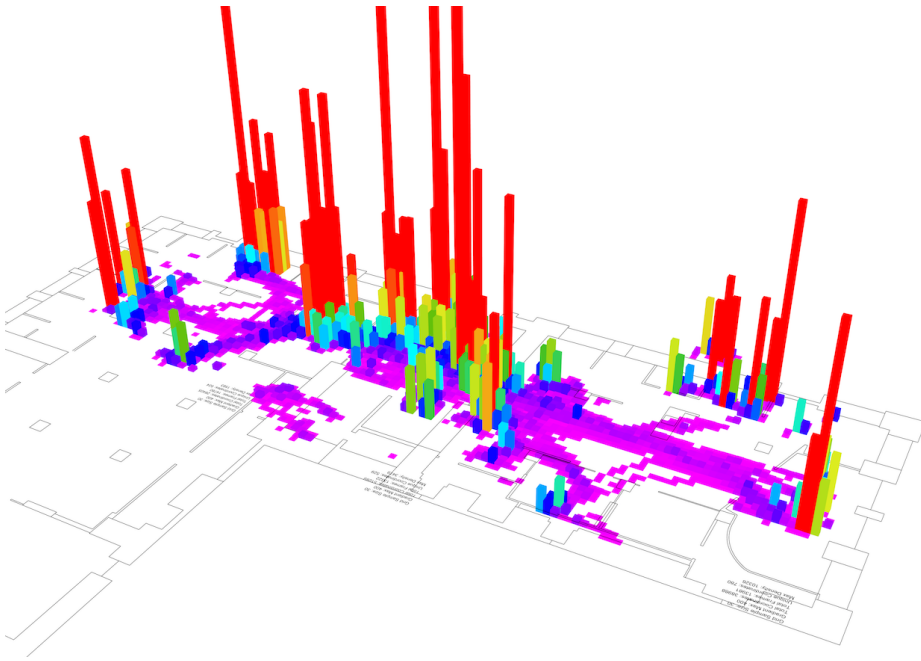


Figure 40.: Analysis of occupancy density visualised as a hybrid between a heat map and a 3D bar graph.

which proves relevant not only for spatial analysis but equally for understanding where the object's generic interface reveals pressures at certain times of the day and how the spatial arrangements become either overloaded or underused.

Part VI

DISCUSSION

9

CONCLUSION

This thesis presents a novel approach for person authentication in smart environments that improves the resilience and security of facilities. The main contributions of the thesis are the (1) development and evaluation of persistent authentication, (2) introduction of remote biometrics for continuous user authentication, (3) formulation of error-rate-based fusion of biometric systems, and (4) implementation of data exploration techniques for identification of contextual elements.

1. Persistent authentication institutes a new approach to authentication that combines traditional access control systems with the sensing technologies and tracking capabilities offered by smart environments. Users are tracked from the point of initial authentication to the point where authorisation is requested by location-based services. The result is a calm approach to authentication, where users are transparently authenticated towards the system. Persistent authentication enables the secure provision of location-based services and implies a shift in the current authentication paradigm from a single discrete event to a continuous session.

2. Remote biometrics are introduced to facilitate the continuous authentication of users, by periodically verifying the identity of each tracked principal. Facial recognition and appearance analysis are used to form a multi-factor biometric authentication approach that increases the reliability of authentication and allows the persistent authentication system to re-associate lost tracking sessions.

3. Error-rate-based fusion is presented as a novel technique to fuse the scores of multiple biometric systems. In error-rate-based fusion the individual biometric scores are transformed into objective evidences and fused using Bayesian inference. This solves a common problem that occurs in sensor fusion, namely, that the evaluation of multiple biometric characteristics produce results that are incompatible, due to different score ranges and different probability distributions.

4. The contextual and behavioural patterns of the occupants are identified to facilitate the detection of insiders, intruders, and hostile reconnaissance by the persistent authentication system. A dedicated tool is developed to process and structure the very large quantities of data generated by the surveillance system. The learned patterns are considered in combination with the spatial configuration of the facility and the correlation between the service functions provided in the facility and the patterns in movement are established. The findings are integrated with state of the art data visualisation to present the information in a clear and useful manner.

The persistent authentication prototype is evaluated with regard to the persistence, robustness and scalability to assess the performance in varying environmental conditions, using different smart technologies. The evaluation consists of a combination of case studies, scenario-based validations and evaluation on public-domain datasets. The evaluation shows that the persistent authentication prototype provides good results, even with low quality video data and no prior information about the principals. More specifically, the persistent authentication prototype achieves 91.78% accuracy on the Lisbon CAVIAR dataset.

In addition, the data provided by the persistent authentication system enables the analysis of occupancy density. This analysis reveals a strong correlation between the movement of occupants and the spatial integration of services in a facility. This relation shows how the spatial arrangements become either overloaded or underused during the day and is of vital importance when it comes to evacuation or emergency situations. Further, the analysis is important in the identification of choke points and high density areas, that are prime targets for attacks and hostile reconnaissance.

Finally, we conclude that persistent authentication offers an effective integrated protection measure that is distributed directly in the facility and is non-intrusive to the public and affordable to the facility owners. Persistent authentication is suitable for security sensitive applications and can help protect the facility against insiders and intruders who seek to cause disruption, terror and other types of crime.

9.1 FUTURE WORK

The implementation of persistent authentication as part of a protection measure for counter-terrorism is a truly multi-disciplinary effort, that touches many research areas. As the presence of technology progressively increases to pervade our urban environments, it is clear that future work lies in pervasive computing, supported by areas such as computer vision and anomaly detection.

These disciplines are all still young, but with much potential driven by a growing interest in surveillance applications and the availability of sensors and processors at reasonable costs. Further, the increasing maturity of algorithms and processing techniques makes previously unthinkable problems solvable in real-time. However, as Szeliski et al. [173] so eloquently state: *"It may be many years before computers can name and outline all of the objects in a photograph with the same skill as a two year old child."*.

The immediate directions for future work are in testing and evaluating the persistent authentication prototype on larger datasets with more complex camera and sensor setups. In addition, the following problems are worth exploring in future work:

- Including additional biometric characteristics, such as gait analysis or long range iris recognition. Further, improvements to the accuracy of the implemented biometrics, to help overcome errors caused by inconsistent lighting or varying poses of the principals.
- Experimenting with different segmentation and feature extraction methods to improve the reliability of the labelling, especially for crowded scenes. Also, investigating scene invariant feature extraction and auto calibration methods for more accurate scene adjustments.
- Integrating recent advancements from the state of the art in person tracking into the persistent authentication tracker. Further, investigating additional options for initialising and terminating tracks between multiple cameras.

Finally, persistent authentication provides important functionality that, besides advanced security applications, are attractive to many other areas. Future applications could include health care, multimedia and home automation, warehousing and applications in business intelligence.

Part VII

APPENDIX

A

COMPARATIVE EVALUATION

For comparative purposes with regard to previous work, we evaluate the accuracy of the contextual awareness provided by different smart environments and the impact of using different sensing technologies. We employ the same metrics and test cases as presented by Kirschmeyer et al. [5] in their evaluation of the original *PAISE* prototype. The implementation of the prototype is evaluated on the basis of its persistence, robustness, and scalability in various conditions, including both expected use, usurpation and malicious behaviour.

A.1 PERSISTENCE

The test of persistence addresses the system's ability to maintain the functional requirements of tracking. The tests are devised to evaluate the system's capacity to track principals and preserve authentication sessions in scenes with multiple mobile principals that are interacting with each other and creating occlusions. The objective of the tracking algorithm is to accurately track principals in all the tests, meaning that neither interactions between the principals nor occlusions should affect the system's performance. The results of the persistence evaluation is shown in Table 12.

The first three experiments validate that tracking works under normal conditions with no interactions. Experiments 4, 5 and 6 show that the system is able to handle partial occlusions and interaction between the principals. The last experiment shows the differences between the sensing technologies, as the two principals are successfully tracked with the time-of-flight camera and not with the CCTV cameras. The depth information provided by the time-of-flight cameras gives a better segmentation of the scene in cases with heavy occlusions and allows the system to track both principals. For the CCTV cameras, the image segmentation process is unable distinguish the two principals during the embrace, and consequently the system needs to rely on the spatio-temporal analysis and continuous authentication to re-associate the authentication sessions with the cor-

Table 12.: The results of the persistence evaluation

Test	Scenario	CCTV	TOF
1	Two principals walk in the same direction.	✓	✓
2	Two principals walk in opposite directions.	✓	✓
3	Two principals cross each other's paths.	✓	✓
4	Two principals bump into each other while walking.	✓	✓
5	Two principals talk. Minimal occlusion of one principal.	✓	✓
6	Two principals shake hands. Moderate occlusion of one principal.	✓	✓
7	Two principals embrace each other warmly. Heavy occlusion of both principal.	(✓)	✓

rect principals. As a result, the success or failure in this case is determined by the availability of the biometrics or the motion consistency of the scene.

We note that the loss of authentication sessions does not correspond to a security vulnerability; it simply requires the principals to re-authenticate before they can make use of further location-based services.

A.2 ROBUSTNESS

The test of robustness addresses the system's ability to function outside the normal operating conditions. The tests include changes in the illumination of the scene and change in the velocity and appearance of the principals. The scene is primarily illuminated with ceiling mounted fluorescent lights. However, large windows allow direct sunlight to pour in, which can drastically change the illumination of the scene.

Furthermore, the system's ability to resist malicious attempts of manipulation, such as usurping the identity of legitimate users or accessing restricted areas without proper authorisation are evaluated. The results of the robustness tests are shown in Table 13.

The two first experiments evaluate the system's ability to preserve authentication sessions when the illumination of the scene changes, either periodically during the day, or drastically, for instance caused by direct sunlight or fluorescent lights that are turned on/off. These changes affect the background model of the scene and may cause labelling ambiguities, however the Gaussian Mixture Model is a dynamic method and is sufficiently adaptive to handle even drastic changes in the illumination.

Experiment 3, 4 and 5 evaluates the system's ability to successfully track principals despite change in their appearance or velocity. The velocity of principals greatly impacts the time-of-flight setup, which is unable to track principals that either move very fast or make quick changes in their direction. The hardware is simply not fast enough to register the principals in these cases. Conversely, the CCTV cameras are unaffected by the velocity of the principals, and track principals in test 4 and 5 without problems.

Experiment 6, 7 and 8 verifies that the authentication and authorisation zones are correctly defined, such that only a single principal at a time is allowed access

Table 13.: The results of the robustness evaluation

Test	Scenario	CCTV	TOF
1	The illumination of the scene changes periodically during the day.	✓	✓
2	Direct sunlight causes a drastic change in the illumination of the scene.	✓	✓
3	A principal changes clothes, causing a drastic change in appearance.	✓	✓
4	A principal makes a quick dash forward.	✓	✗
5	A principal runs at high speed.	✓	✗
6	Illegitimate access is attempted by an unauthenticated principal.	✓	✓
7	An usurper tries to tailgate a legitimate principal through the authorisation zone.	✓	✓
8	Two principals try to authenticate at the authentication zone with only a single smart-card.	✓	✓
9	Denial of Service is attempted as a principal points an infrared light at the camera.	✓	✗

and that the resolution of the sensors is sufficient to determine when multiple principals are in the zones. Additionally, the experiment verifies that the security policy protecting the restricted area is correctly enforced by the system.

In the last experiment it is possible for a principal to execute a denial of service attack on the system by blinding the time-of-flight camera with an infrared light source. The blinded camera is unable to track principals in the scene and all current authentication sessions are lost. The CCTV cameras have an infrared cut-off filter that blocks infrared wavelength, thus most of the effect of the attack is mitigated. In either case the attacker has to be very close to the cameras, making the applications of the attack restricted by the placement of the cameras.

The evaluation shows that the robustness of the persistent authentication system is high, and that the system is suitable for security sensitive purposes. Moreover, we conjecture that the persistence and robustness can be improved significantly by using a multi-modal sensor environment, e.g., by correlating the results from multiple sensors covering the same scene. Ideally, a multi-model system integrates the time-of-flight and CCTV solutions, such that the strengths of both sensor types are utilised.

A.3 SCALABILITY

The scalability of the system depends on the prototype's ability to track and authenticate a large number of principals in an extensive environment. In addition, the scalability depends on the effort required to install and maintain the system, i.e., the expenses incurred by a large hardware installation including many sensors, authentication points and similar smart technology. Our experience with installing and maintaining the system during the case studies and evaluations suggests that the scalability in this regard is sufficient for even very extensive environments.

The following experiments test the system's performance with a varying number of mobile and stationary principals, moving and queueing in the premise. Figure 41 shows three examples of the varying number of principals occupying the scene, in the figure each dot represents a principal and the colour of the dot indicates if the principal is either moving (red) or static (blue). The results of the evaluation are summarised in Table 14

Table 14.: The results of the scalability evaluation

Test	Scenario	CCTV	TOF
1	Low density, with large unoccupied areas.	✓	✓
2	Average density, with queues in some areas and interaction between principals.	✓	(✓)
3	High density, with many interactions between mobile principals and compact queues.	✗	✗

The tests show that for low population densities, the performance of the system with both sensor technologies are good. When queues start to form and the scene gets crowded, the CCTV solution is still performing satisfactory, whereas, the time-of-flight solution have difficulties tracking all principals. This is caused partly by the limitations of the hardware, and partly due to the mounting angle of the camera, which creates more occlusions of the principals than the ceiling mounted CCTV cameras.

In the third experiment, the scene is considered very crowded, with queues in many areas. The time-of-flight solution is unable to distinguish individual principals due to the many occlusions. The CCTV solution can track moving principals, but are unable maintaining authentication sessions of principals in the queues. The segmentation of these principals contain a lot of noise, thus the labelling used for tracking are ambiguous.

The overall the scalability of the system is considered adequate. Based on these tests and on our previous experiences with installing the persistent authentication system, we recommend CCTV cameras for any larger installation. The performance is higher and the cameras are considerably cheaper.

Figure 41.: The varying number of principals (dots). Low density (top), average density (middle) and high density (bottom).

BIBLIOGRAPHY

[1] M. Reiter and P. Rohatgi, "Homeland Security," *IEEE Internet Computing*, vol. 8, pp. 16–17, Nov. 2004.

[2] M. Weiser and J. Brown, "Designing calm technology," *PowerGrid Journal*, pp. 1–5, 1996.

[3] M. Weiser, "The Computer for the 21st Century," *Scientific American*, 1991.

[4] D. Cook and S. Das, *Smart Environments: Technology, Protocols and Applications*. Wiley, 2004.

[5] M. Kirschmeyer and M. S. Hansen, "Persistent Authentication in Smart Environments," *Technical University of Denmark, IMM-THESIS: 2008-16*, 2008.

[6] H. Lieberman and T. Selker, "Out of context: Computer systems that adapt to, and learn from, context," *IBM Systems Journal*, vol. 39, pp. 617–632, 2000.

[7] B. Schilit, N. Adams, and R. Want, "Context-aware computing applications," in *Workshop on Mobile Computing Systems and Applications*, pp. 85–90, IEEE Comput. Soc. Press, 1994.

[8] M. Argumosa, *Development of Face Recognition: Infancy to Early Childhood*. PhD thesis, Florida International University, 2010.

[9] S. Kouider, C. Stahlhut, S. V. Gelskov, L. S. Barbosa, M. Dutat, V. De Gardelle, A. Christophe, S. Dehaene, and G. Dehaene-Lambertz, "A neural marker of perceptual consciousness in infants.," *Science*, vol. 340, pp. 376–380, 2013.

[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Computer Vision and Pattern Recognition*, 2001.

[11] M. Jones and P. Viola, "Robust real-time object detection," *Workshop on Statistical and Computational Theories of Vision*, 2001.

[12] F. Council, "Authentication in an electronic banking environment," 2001.

[13] L. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication," *Proceedings of the IEEE*, vol. 91, pp. 2021–2040, Dec. 2003.

[14] F. Council, "Frequently Asked Questions on FFIEC Guidance on Authentication in an Internet Banking Environment," Jan. 2006.

[15] J. Bardram, R. Kjæ r, and M. Pedersen, "Context-aware user authentication–supporting proximity-based login in pervasive computing," *UbiComp 2003: Ubiquitous Computing*, 2003.

[16] H. Christensen and J. Bardram, "Supporting human activities—exploring activity-centered computing," *UbiComp 2002: Ubiquitous Computing*, 2002.

[17] J. Bardram, "Plans as situated action: an activity theory approach to workflow systems," *ECSCW'97 Proceedings of the fifth conference on European Conference on Computer-Supported Cooperative Work*, pp. 17 – 32, 1997.

[18] J. Bardram, "Designing for the dynamics of cooperative work activities," *Proceedings of the 1998 ACM conference on Computer supported cooperative work - CSCW '98*, pp. 89–98, 1998.

[19] D. Denning and P. MacDoran, "Location Based authentication: grounding cyberspace for better security," *Computer Fraud & Security*, 1996.

[20] M. Corner and B. Noble, "Zero-interaction authentication," *Proceedings of the 8th annual international Conference on Mobile Computing and Networking*, p. 1, 2002.

[21] B. D. Noble and M. D. Corner, "The case for transient authentication," *Proceedings of the 10th workshop on ACM SIGOPS European workshop: beyond the PC - EW10*, p. 24, 2002.

[22] M. Corner and B. Noble, "Protecting applications with transient authentication," *International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2003.

[23] L. Bussard and Y. Roudier, "Embedding distance-bounding protocols within intuitive interactions," *Security in Pervasive Computing*, pp. 143–156, 2004.

[24] S. Capkun and J. Hubaux, "Secure positioning of wireless devices with application to sensor networks," *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, 2005.

[25] S. Capkun and J. Hubaux, "Secure positioning in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 221–232, Feb. 2006.

[26] T. Kindberg, K. Zhang, and N. Shankar, "Context authentication using constrained channels," *Mobile Computing Systems and Applications*, 2002.

[27] H. Aubert, "RFID technology for human implant devices," *Comptes Rendus Physique*, vol. 12, pp. 675–683, Sept. 2011.

[28] M. Tistarelli, S. Z. Li, and R. Chellappa, "Handbook of Remote Biometrics," Mar. 2009.

[29] A. Klosterman and G. Ganger, "Secure Continuous Biometric-Enhanced Authentication," tech. rep., Parallel Data Laboratory, 2000.

[30] D. Bhattacharyya, "Biometric authentication: A review," *International Journal of u-and e-Service*, vol. 2, no. 3, pp. 13–28, 2009.

[31] S. Cole, "More than zero: Accounting for error in latent fingerprint identification," *The Journal of Criminal Law and Criminology (1973-)*, vol. 95, no. 3, pp. 985–1078, 2005.

[32] J. T. Committee, "ISO/IEC TR 24722:2007 Biometrics - Multimodal and other multibiometric fusion," tech. rep., ISO/IEC JTC 1/SC 37, 2013.

[33] T. Sim and S. Zhang, "Continuous Verification Using Multimodal Biometrics," *Pattern Analysis and Machine Intelligence*, pp. 562–570, 2007.

[34] J. Muncaster and M. Turk, "Continuous multimodal authentication using dynamic Bayesian networks," *Proceedings of the 2nd Workshop of Multimodal User Authentication*, 2006.

[35] A. Altinok and M. Turk, "Temporal integration for continuous multimodal biometrics," *Proceedings of the Workshop on Multimodal User Authentication*, no. 1, 2003.

[36] K. Niinuma, U. Park, and A. K. Jain, "Soft Biometric Traits for Continuous User Authentication," *IEEE Transactions on Information Forensics and Security*, vol. 5, pp. 771–780, Dec. 2010.

[37] P. Norman, *Multi-system Biometric Authentication: Optimal Fusion and User-Specific Information*. PhD thesis, Swiss Federal Institute of Technology in Lausanne (EPFL), 2006.

[38] Q. Tao and R. Veldhuis, "Threshold-optimized decision-level fusion and its application to biometrics," *Pattern Recognition*, vol. 42, no. 5, pp. 823–836, 2009.

[39] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain, "Likelihood ratio-based biometric score fusion.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 342–347, 2008.

[40] J. Kittler, "Combining classifiers: A theoretical framework," *Pattern Analysis and Applications*, vol. 1, pp. 18–27, Mar. 1998.

[41] J. Kittler and M. Hatef, "On combining classifiers," *Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[42] J. Kittler and F. M. Alkoot, "Sum versus vote fusion in multiple classifier systems," *Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 110–115, 2003.

[43] K. Kryszczuk and J. Richiardi, "Error handling in multimodal biometric systems using reliability measures," *Proc. 12th European Conference on Signal Processing*, pp. 0–3, 2005.

[44] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun, "Discriminative multimodal biometric authentication based on quality measures," *Pattern Recognition*, vol. 38, pp. 777–779, May 2005.

[45] K. Nandakumar, a.K. Jain, and S. Dass, "Quality-based Score Level Fusion in Multibiometric Systems," *18th International Conference on Pattern Recognition (ICPR'06)*, pp. 473–476, 2006.

[46] D. E. Maurer and J. P. Baker, "Fusing multimodal biometrics with quality estimates via a Bayesian belief network," *Pattern Recognition*, vol. 41, pp. 821–832, Mar. 2008.

[47] K. A. Toh, W. Yau, E. Lim, L. Chen, and C.-h. Ng, "Fusion of auxiliary information for multi-modal biometrics authentication," *Biometric Authentication*, pp. 678–685, 2004.

[48] K. A. Toh, W.-Y. Yau, and X. Jiang, "A reduced multivariate polynomial model for multimodal biometrics and classifiers fusion," 2004.

[49] K. A. Toh and W.-Y. Yau, "Combination of hyperbolic functions for multimodal biometrics data fusion," *IEEE transactions on systems man and cybernetics Part B Cybernetics a publication of the IEEE Systems Man and Cybernetics Society*, vol. 34, no. 2, pp. 1196–1209, 2004.

[50] K. A. Toh, "Error-rate based biometrics fusion," *Advances in Biometrics*, pp. 191–200, 2007.

[51] K. A. Toh, J. Kim, and S. Lee, "Biometric scores fusion based on total error rate minimization," *Pattern Recognition*, vol. 41, pp. 1066–1082, Mar. 2008.

[52] M. Hanmandlu and J. Grover, "Error Level Fusion of Multimodal Biometrics," *Journal of Pattern Recognition*, vol. 2, pp. 278–297, 2011.

[53] Y. Li, J. Yin, J. Long, and E. Zhu, "A Novel Method for Multibiometric Fusion Based on FAR and FRR," *Modeling Decisions for Artificial Intelligence*, pp. 194–204, 2009.

[54] S. Bengio, C. Marcel, S. Marcel, and J. Mariethoz, "Confidence Measures for Multimodal Identity Verification," *Information Fusion*, vol. 3, no. 4, pp. 267–276, 2002.

[55] N. Poh and S. Bengio, "Improving Fusion with Margin-Derived Confidence In Biometric Authentication Tasks," *Lecture Notes in Computer Science*, vol. 3546, pp. 347–356, 2005.

[56] N. Poh and S. Bengio, "A Novel Approach to Combining Client-Dependent and Confidence Information in Multimodal Biometricactive," *Lecture Notes in Computer Science*, vol. 3546, p. 1120, 2005.

[57] V. Chatzis, A. G. Bors, and I. Pitas, "Multimodal decision-level fusion for person authentication," *IEEE Transactions on Systems Man and Cybernetics Part A Systems and Humans*, vol. 29, no. 6, pp. 674–680, 1999.

[58] E. Bigün, J. Bigün, B. Duc, and S. Fischer, "Expert conciliation for multi modal person authentication systems by Bayesian statistics," *Biometric Person Authentication*, 1997.

[59] J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Multimodal biometric authentication using quality signals in mobile communications," 2003.

[60] E. S. Bigün, "Risk analysis of catastrophes using experts' judgements: An empirical study on risk analysis of major civil aircraft accidents in Europe," *European Journal Of Operational Research*, vol. 87, no. 3, pp. 599–612, 1995.

[61] H. Fronthaler, K. Kollreider, J. Bigun, J. Fierrez, F. Alonso-Fernandez, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Fingerprint Image-Quality Estimation and its Application to Multialgorithm Verification," 2008.

[62] N. Poh and J. Kittler, "A unified framework for biometric expert fusion incorporating quality measures," *Pattern Analysis and Machine Intelligence*, pp. 1–14, 2012.

[63] R. N. Rodrigues, L. L. Ling, and V. Govindaraju, "Robustness of multimodal biometric fusion methods against spoof attacks," *Journal of Visual Languages & Computing*, vol. 20, pp. 169–179, June 2009.

[64] Z. Kalal, *Tracking-learning-detection*. PhD thesis, University of Surrey, 2010.

[65] H. Liu and H. Darabi, "Survey of wireless indoor positioning techniques and systems," *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 6, pp. 1067–1080, 2007.

[66] R. Mautz and S. Tilch, "Survey of optical indoor positioning systems," *2011 International Conference on Indoor Positioning and Indoor Navigation*, pp. 1–7, Sept. 2011.

[67] T. Wilhelm, H. Böhme, and H. Gross, "Sensor fusion for vision and sonar based people tracking on a mobile service robot," *Proc. Int. Workshop on Dynamic Perception*, pp. 315–320, 2002.

[68] P. Srinivasan and D. Birchfield, "Design of a pressure sensitive floor for multimodal sensing," *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, 2005.

[69] C. Drane, "Positioning GSM Phones," *Communications Magazine*, 1998.

[70] B. Fang, "Simple solution for hyperbolic and related position fixes," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 26, no. 5, 1990.

[71] D. Torrieri, "Statistical theory of passive location systems," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 00, no. 2, pp. 183–198, 1984.

[72] J. Ng, S. Chan, and K. Kan, "Providing location estimation within a metropolitan area based on a mobile phone network," *Database and Expert Systems Applications, 2002. Proceedings. 13th International Workshop on*, 2002.

[73] K. Chu and J. Ng, "Providing Location Services within a Radio Cellular Network Using Ellipse Propagation Model," *19th International Conference on Advanced Information Networking and Applications (AINA'05) Volume 1 (AINA papers)*, vol. 1, pp. 559–564, 2005.

[74] R. Want, A. Hopper, V. Falcão, and J. Gibbons, "The active badge location system," *ACM Transactions on Information Systems (TOIS)*, 1992.

[75] P. Bahl and V. Padmanabhan, "RADAR: An In-Building RF-based User Location and Tracking System," *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings*, 2000.

[76] J. Hightower, R. Want, and G. Borriello, "SpotON: An indoor 3D location sensing technology based on RF signal strength," *UW CSE 00-02-02, University of Washington, Department of Computer Science and Engineering, Seattle*, 2000.

[77] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: Indoor Location Sensing Using Active RFID," *Wireless Networks*, vol. 10, pp. 701–710, Nov. 2004.

[78] S. Feldmann, "An Indoor Bluetooth-Based Positioning System: Concept, Implementation and Experimental Evaluation.," *International Conference on Wireless Networks*, 2003.

[79] M. Delafontaine, M. Versichele, T. Neutens, and N. Van de Weghe, "Analysing spatiotemporal sequences in Bluetooth tracking data," *Applied Geography*, vol. 34, pp. 659–668, May 2012.

[80] E. A. Fry and L. A. Lenert, "MASCAL: RFID tracking of patients, staff and equipment to enhance hospital response to mass casualty events.," *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, pp. 261–265, 2005.

[81] E. Trucco and K. Plakas, "Video Tracking: A Concise Survey," *IEEE Journal of Oceanic Engineering*, vol. 31, pp. 520–529, Apr. 2006.

[82] T. Bouwmans, "Recent advanced statistical background modeling for foreground detection-a systematic survey," *Recent Patents on Computer Science*, no. 33, 2011.

[83] B. Lee and M. Hedley, "Background Estimation for Video Surveillance," in *Image & Vision Computing New Zealand (IVCNZ '02)*, pp. 315–320, 2002.

[84] N. J. B. McFarlane and C. P. Schofield, "Segmentation and tracking of piglets in images," *Machine Vision and Applications*, vol. 8, no. 3, pp. 187–193, 1995.

[85] J. Zheng and Y. Wang, "Extracting Roadway Background Image: a Mode-Based Approach," *Journal of the Transportation Research Board*, no. 206, pp. 1–15, 2006.

[86] C. Wren and A. Azarbayejani, "Pfinder: Real-Time Tracking of the Human Body," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 780–785, 1997.

[87] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*, pp. 246–252, 1999.

[88] C. Stauffer and W. W. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.

[89] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," *Computer Vision—ECCV 2000*, 2000.

[90] M. Sigari, N. Mozayani, and H. Pourreza, "Fuzzy running average and fuzzy background subtraction: concepts and application," *International Journal of Computer Science and Network Security*, vol. 8, no. 2, pp. 138–143, 2008.

[91] F. E. Baf, T. Bouwmans, and B. Vachon, "Fuzzy integral for moving object detection," *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World*

Congress on Computational Intelligence). IEEE International Conference on, vol. 1, pp. 1729–1736, June 2008.

[92] F. E. Baf, T. Bouwmans, and B. Vachon, "Type-2 fuzzy mixture of Gaussians model: application to background modeling," *Advances in Visual Computing*, 2008.

[93] D. E. Butler, V. M. Bove, and S. Sridharan, "Real-Time Adaptive Foreground/Background Segmentation," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 14, pp. 2292–2304, 2005.

[94] T. Gao, Z. Liu, S. Yue, J. Zhang, J. Mei, and W. Gao, "Robust background subtraction in traffic video sequence," *Journal of Central South University of Technology*, pp. 187–195, 2010.

[95] R. Kalman, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, 1960.

[96] N. Friedman and S. Russell, "Image segmentation in video sequences: A probabilistic approach," *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, pp. 175–181, 1997.

[97] S. Brutzer, B. Hoferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," *Cvpr 2011*, pp. 1937–1944, June 2011.

[98] D. Parks, "Evaluation of Background Subtraction Algorithms with Post-Processing," *Advanced Video and Signal Based Surveillance*, 2007.

[99] R. Cucchiara and C. Grana, "Detecting moving objects, ghosts, and shadows in video streams," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 10, pp. 1337–1342, 2003.

[100] N. Oliver, "A Bayesian Computer Vision System for Modeling Human Interactions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 831–843, 2000.

[101] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision.," *IJCAI*, vol. 130, pp. 121–130, 1981.

[102] J. Sánchez Pérez, E. Meinhardt-Llopis, and G. Facciolo, "TV-L1 Optical Flow Estimation," *Image Processing On Line*, vol. 2013, pp. 137–150, July 2013.

[103] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, Aug. 1981.

[104] G. Farneback, "Fast and accurate motion estimation using orientation tensors and parametric motion models," *Pattern Recognition*, 2000.

[105] G. Farneback, "Very high accuracy velocity estimation using orientation tensors, parametric motion, and simultaneous segmentation of the motion field," *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 2001.

[106] G. Welch and G. Bishop, "An introduction to the Kalman filter," tech. rep., University of North Carolina, 1995.

[107] A. Lipton, H. Fujiyoshi, and R. Patil, "Moving target classification and tracking from real-time video," *Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on*, 1998.

[108] P. S. Maybeck, "Stochastic models, estimation, and control," 1979.

[109] S.-K. Weng, C.-M. Kuo, and S.-K. Tu, "Video object tracking using adaptive Kalman filter," *Journal of Visual Communication and Image Representation*, vol. 17, pp. 1190–1208, Dec. 2006.

[110] R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," *Journal of Basic Engineering*, vol. 83, no. 1, p. 95, 1961.

[111] S. Julier and J. Uhlmann, "New extension of the Kalman filter to nonlinear systems," *AeroSense'97*, 1997.

[112] B. Saulson and K. Chang, "Comparison of nonlinear estimation for ballistic missile tracking," *AeroSense 2003*, vol. 5096, pp. 13–24, 2003.

[113] J. Gross, Y. Gu, and M. Rhudy, "Flight-Test Evaluation of Sensor Fusion Algorithms for Attitude Estimation," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 3, 2012.

[114] M. Rhudy, Y. Gu, and M. R., "An Analytical Approach for Comparing Linearization Methods in EKF and UKF," *International Journal of Advanced Robotic Systems*, p. 1, 2013.

[115] F. Orderud, "Comparison of kalman filter estimation approaches for state space models with nonlinear measurements," *Proceedings of scandinavian conference on simulation and modeling*, no. 7491, 2005.

[116] M. Isard and A. Blake, "CONDENSATION — Conditional Density Propagation for Visual Tracking," *International journal of computer vision*, vol. 29, no. 1, pp. 5–28, 1998.

[117] M. Breitenstein and F. Reichlin, "Robust tracking-by-detection using a detector confidence particle filter," *Computer Vision, 2009 IEEE 12th International Conference on*, no. Iccv, 2009.

[118] S. Lee and M. West, "Performance Comparison of the Distributed Extended Kalman Filter and Markov Chain Distributed Particle Filter (MCDPF)," *Proc. 2nd IFAC Workshop on Distributed Estimation and Control in Networked Systems (NecSys' 10)*, pp. 2–7, 2010.

[119] S. Won, W. Melek, and F. Golnaraghi, "A Kalman/particle filter-based position and orientation estimation method using a position sensor/inertial measurement unit hybrid system," *Industrial Electronics, IEEE Transactions on*, vol. 57, no. 5, pp. 1787–1798, 2010.

[120] J. Shi and C. Tomasi, "Good features to track," *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, 1994.

[121] S. Birchfield, "KLT: An Implementation of the Kanade-Lucas-Tomasi Feature Tracker," 2007.

[122] G. Takacs, V. Chandrasekhar, S. Tsai, D. Chen, R. Grzeszczuk, and B. Girod, "Unified Real-Time Tracking and Recognition with Rotation-Invariant Fast Features," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 934–941, June 2010.

[123] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on Information Theory*, vol. 21, no. 1, 1975.

[124] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, pp. 1–37, 2002.

[125] A. Elgammal, R. Duraiswami, and L. Davis, "Efficient kernel density estimation using the fast gauss transform with applications to color modeling and tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1499–1504, Nov. 2003.

[126] R. Duraiswami and L. Davis, "Efficient Mean-Shift Tracking via a New Similarity Measure," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 176–183, 2005.

[127] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-Learning-Detection.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 6, pp. 1–14, Dec. 2011.

[128] S. J. Prince, *Computer Vision. Models, Learning, and Inference.* Cambridge, 2012.

[129] C. Bettini and S. Jajodia, "Privacy in location-based applications: research issues and emerging trends," *New York*, vol. 5599, no. 3, p. 224, 2009.

[130] U. Leonhardt and J. Magee, "Towards a general location service for mobile environments," in *Proceedings*, pp. 43–50, 1996.

[131] C. D. Jensen, K. Geneser, and I. Willemoes-Wissing, "Sensor Enhanced Access Control: Extending Traditional Access Control Models with Context-Awareness," *Trust Management VII*, pp. 1–16, 2013.

[132] A. Kapadia, T. Henderson, J. J. Fielding, and D. Kotz, "Virtual Walls : Protecting Digital Privacy in Pervasive Environments," in *PERVASIVE'07 Proceedings of the 5th international conference on Pervasive computing*, pp. 162–179, 2007.

[133] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys (CSUR)*, 2009.

[134] R. Boguslaw and A. F. Westin, "Privacy and Freedom.," 1968.

[135] P. K. Nayar, "Human rights in the global information society," *Journal of the American Society for Information Science & Technology*, vol. 58, no. 14, pp. 2407–2408, 2007.

[136] D. Banisar and S. Davies, "Global Trends in Privacy Protection: An International Survey of Privacy, Data Protection, and Surveillance Laws and Developments," *J. Marshall J. Computer & Info. L.*, 1999.

[137] N. Poh, T. Bourlai, J. Kittler, L. Allano, F. Alonso-Fernandez, O. Ambekar, J. Baker, B. Dorizzi, O. Fatukasi, J. Fierrez, H. Ganster, J. Ortega-Garcia, D. Maurer, A. A. Salah, T. Scheidat, and C. Vielhauer, "Benchmarking Quality-Dependent and Cost-Sensitive Score-Level Multimodal Biometric Fusion Algorithms," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 849–866, 2009.

[138] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, no. 2, pp. 28–31 Vol.2, 2004.

[139] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, pp. 773–780, May 2006.

[140] B. White and M. Shah, "Automatically Tuning Background Subtraction Parameters using Particle Swarm Optimization," *In Multimedia and Expo, 2007 IEEE International Conference on (2007)*, 2007.

[141] M. Shah, J. D. Deng, and B. J. Woodford, "Improving Mixture of Gaussians background model through adaptive learning and Spatio-Temporal voting," *2013 IEEE International Conference on Image Processing*, pp. 3436–3440, Sept. 2013.

[142] S. Suzuki, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 46, pp. 32–46, 1985.

[143] J. Bouguet, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel Corporation*, vol. 1, no. 2, pp. 1–9, 2001.

[144] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, 1991.

[145] R. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Human Genetics*, 1936.

[146] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711–720, July 1997.

[147] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 800, 1997.

[148] E. Anderson, "The Irises of the Gaspe Peninsula," *Bulletin of the American Iris Society*, vol. 59, pp. 2–5, 1935.

[149] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Wiley, 2001.

[150] M. Flickner, H. Sawhney, and W. Niblack, "Query by image and video content: the QBIC system," *Computer*, 1995.

[151] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern Recognition*, vol. 40, pp. 1106–1122, Mar. 2007.

[152] D. Hall, F. Pélisson, O. Riff, and J. Crowley, "Brand identification using Gaussian derivative histograms," *Machine Vision and Applications*, 2004.

[153] A. Jain, S. Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," *Biometric Authentication*, no. July, pp. 1–7, 2004.

[154] P. Mazzeo and L. Giove, "HSV and RGB color histograms comparing for objects tracking among non overlapping FOVs, using CBTF," *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, pp. 498–503, 2011.

[155] M. Luber, G. D. Tipaldi, and K. O. Arras, "Better models for people tracking," *2011 IEEE International Conference on Robotics and Automation*, pp. 854–859, May 2011.

[156] M. Mucientes and W. Burgard, "Multiple Hypothesis Tracking of Clusters of People," *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 692–697, Oct. 2006.

[157] M. Li and J. Lavest, "Some aspects of zoom lens camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 11, pp. 1105–1110, 1996.

[158] D. Bailey, "A New Approach to Lens Distortion Correction," *Proceedings Image and Vision Computing New Zealand 2002*, 2002.

[159] K. Gribbon, "A real-time FPGA implementation of a barrel distortion correction algorithm with bilinear interpolation," *Image and Vision Computing New Zealand*, 2003.

[160] R. Szeliski, "Image Alignment and Stitching: A Tutorial," *Foundations and Trends® in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2006.

[161] M. B. Stegmann, B. K. Ersboll, and R. Larsen, "FAME–a flexible appearance modeling environment.," 2003.

[162] E. Tabassi and P. Grother, "NIST Biometric Scores Set," tech. rep., National Institute of Standards and Technology, 2004.

[163] R. Fisher, "CAVIAR Test Case Scenarios," 2004.

[164] S. Dass, K. Nandakumar, and A. Jain, "A Principled Approach to Score Level Fusion in Multimodal Biometric Systems," *Audio-and Video-Based Biometric ...*, no. i, 2005.

[165] N. Srinivas, "Fusing correlated data from multiple classifiers for improved biometric verification," *Information Fusion, 2009. FUSION '09. 12th International Conference on*, pp. 1504–1511, 2009.

[166] O. Fatukasi, J. Kittler, and N. Poh, "Quality controlled multimodal fusion of biometric experts," *Progress in Pattern Recognition, Image Analysis and Applications*, 2007.

[167] M. He, S.-J. Horng, P. Fan, R.-S. Run, R.-J. Chen, J.-L. Lai, M. K. Khan, and K. O. Sentosa, "Performance evaluation of score level fusion in multimodal biometric systems," *Pattern Recognition*, vol. 43, pp. 1789–1800, May 2010.

[168] B. Ulery, A. Hicklin, P. Hallinan, C. Watson, and W. Fellner, "Studies of Biometric Fusion," tech. rep., The National Institute of Standards and Technology (NIST), 2006.

[169] S. Some, "Use cases based requirements validation with scenarios," *Requirements Engineering, 2005. Proceedings. 13th IEEE International Conference on*, pp. 0–1, 2005.

[170] T. Aven, *Risk analysis: Assessing uncertainties beyond expected values and probabilities*. Wiley, 2008.

[171] D. Cornish, "The procedural analysis of offending and its relevance for situational prevention," *Crime prevention studies*, pp. 151–196, 1994.

[172] R. Clarke and G. Newman, *Outsmarting the terrorists*. Wiley, 2006.

[173] R. Szeliski, "Computer Vision : Algorithms and Applications," *Computer*, vol. 5, p. 832, 2010.